UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

Alexandre Diniz da Costa

A tradução por máquina enriquecida semanticamente com frames e papéis qualia

Juiz de Fora
2020

Alexandre Diniz da Costa

A tradução por máquina enriquecida semanticamente com frames e papéis qualia

Tese apresentada ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Linguística. Área de concentração: Linguística.

Orientador: Prof. Dr. Tiago Timponi Torrent

Juiz de Fora
2020

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Costa, Alexandre Diniz da .
A tradução por máquina enriquecida semanticamente com frames e papéis qualia / Alexandre Diniz da Costa. -- 2020.
244 p.

Orientador: Tiago Timponi Torrent
Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2020.

1. Semântica de Frames. 2. Tradução Automática. 3. Estrutura Qualia. 4. FrameNet. 5. Injeção Terminológica. I. Torrent, Tiago Timponi , orient. II. Título.

Alexandre Diniz da Costa

A tradução por máquina enriquecida semanticamente com frames e papéis qualia

Tese apresentada ao programa de Pós-Graduação em Linguística da Faculdade de Letras da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Linguística. Área de concentração: Linguística.

Aprovada em 16 de dezembro de 2020

BANCA EXAMINADORA

Professor Doutor Tiago Timponi Torrent - Orientador

Universidade Federal de Juiz de Fora

Professora Doutora Helena de Medeiros Caseli

UFSCar

Professor Doutor Diego Campos Moussallem

Paderborn University

Professora Doutora Patrícia Fabiane Amaral da Cunha Lacerda

UFJF

Professora Doutora Sandra Aparecida Faria de Almeida

UFJF

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Tiago Timponi Torrent, por acreditar em meu potencial, me incentivar sempre a crescer, abrir novos caminhos, explorar o mundo.

À Faculdade de Letras e ao Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora, por terem aberto suas portas a mim na construção do conhecimento.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos (Processo 88887.185051/2018-00) que me permitiu realizar meus estudos de doutorado sanduíche na Alemanha pelo Programa PROBRAL durante um ano e os investimentos aplicados na pesquisa.

Aos professores Helena de Medeiros Caseli, Diego Campos Moussallem, Patrícia Fabiane Amaral da Cunha Lacerda e Sandra Aparecida de Faria Almeida que aceitaram o convite de participação em minha banca de defesa. Cada colocação contribui substancialmente para meu crescimento pessoal, profissional e para o desenvolvimento desta e de outras pesquisas, bem como para meu amadurecimento acadêmico.

Aos pesquisadores e membros do grupo de Pesquisa FrameNet Brasil, em especial ao Ely, por todas as conversas, dicas, trocas de angústias e experiências, bem como conselhos e tarefas; ao Mairon e à Lívia, por se mostrarem sempre gentis e dispostos a ajudar. Lívia, sua ajuda foi fundamental para que eu conseguisse concluir este trabalho.

À Profa. Dra. Ana Cláudia Peters Salgado por me introduzir na pesquisa acadêmica e sempre estimular na profissão. Foi a peça-chave do meu quebra-cabeça, me auxiliando e orientando mesmo em tempos de “vertigens”.

Aos professores Carolina Alves Magaldi, Rogério de Souza Sérgio Ferreira e Patrícia Fabiane Amaral da Cunha Lacerda que sempre colaboraram para que eu crescesse enquanto estudante, pesquisador e professor.

À Vânia Gomes de Almeida, grande amiga, por dividir comigo as alegrias, tristezas, medos e aventuras, um ano de doutorado-sanduíche juntos, além de me dar conselhos e por orientar diversas vezes acerca de certas decisões a serem tomadas na vida. Obrigado por toda a sua sinceridade, o conhecimento compartilhado e suas dicas. Juntos e cada vez mais fortes.

À minha família, minha mãe, meus irmãos, tios e tias, pelo amor incondicional, admiração, apoio emocional e financeiro.

Ao Fabiano Machado por compartilhar uma vida comigo, com muito amor, apoio, críticas e carinho. Saiba que você estar ao meu lado foi fundamental para meu amadurecimento e crescimento como pessoa e profissional.

“A Cultura

O girino é o peixinho do sapo
O silêncio é o começo do papo
O bigode é a antena do gato
O cavalo é pasto do carrapato

O cabrito é o cordeiro da cabra
O pescoço é a barriga da cobra
O leitão é um porquinho mais novo
A galinha é um pouquinho do ovo

O desejo é o começo do corpo
Engordar é a tarefa do porco
A cegonha é a girafa do ganso
O cachorro é um lobo mais manso

O escuro é a metade da zebra
As raízes são as veias da seiva
O camelo é um cavalo sem sede
Tartaruga por dentro é parede

O potrinho é o bezerro da égua
A batalha é o começo da trégua
Papagaio é um dragão miniatura
Bactérias num meio é cultura.” (ANTUNES, 1996, p. 55).

RESUMO

Esta tese visa a propor uma alternativa de melhoramento semântico à Tradução por Máquina (TM) de domínio específico, através da implementação de dois sistemas híbridos de TM. Para tanto, trabalha-se na incorporação de frames (FILLMORE, 1982) e relações qualia (PUSTEJOVSKY, 1995; PUTEJOVSKY; JEZEK, 2016) como um refinamento semântico disponibilizado a um algoritmo de tradução. A pesquisa se aplica ao domínio específico dos Jogos Olímpicos e envolve o desenvolvimento de um tradutor de sentenças semanticamente melhorado embarcado em uma aplicação computacional (m.knob) em desenvolvimento pela FrameNet Brasil. O aplicativo, um guia de bolso multilíngue, possui três funções principais, sendo (i) um chatbot que recomenda lugares turísticos com base em geolocalização, modelagem de relações qualia e interação com a aplicação, (ii) uma Diciopédia, sendo um repositório multilíngue de frames e termos do turismo e esportes que se estrutura em frames, ULs e relações qualia, e (iii) o tradutor baseado em frames e estrutura qualia. Como metodologia, trabalhamos com a constituição de um corpus de sentenças reais dos esportes, a criação de uma tradução humana de referência e a avaliação através das métricas de tradução por máquina BLEU, TER e HTER do sistema de tradução que representa o estado da arte (S-Base), do sistema de TM por redes neurais com injeção terminológica na etapa de pré-processamento (S-Pré) e do sistema de TM por redes neurais com a injeção terminológica na etapa de pós-edição (S-Pós), sendo os dois últimos desenvolvidos no âmbito desta tese. Os resultados apontam que a injeção terminológica aplicada em sistemas de TM que utilizam redes neurais através do enriquecimento semântico dos sistemas, via implementação de frames e qualia, contribuem para melhorias na geração de traduções mais adequadas para domínios específicos, no caso, o dos esportes.

Palavras-chave: Semântica de Frames. Estrutura Qualia. FrameNet. Tradução Automática. Pré-processamento. Pós-edição.

ABSTRACT

This Ph.D. thesis aims to propose an alternative of semantic improvement to Machine Translation (MT) for specific domains, through the implementation of two hybrid systems of MT. To this end, we work on the integration of frames (FILLMORE, 1982) and qualia relations (PUSTEJOVSKY, 1995; PUTEJOVSKY; JEZEK, 2016) as a semantic refinement approach for a translation algorithm. The piece of research applies to the specific domain of the Olympic Games. It also involves the development of a semantically enriched sentence translator embedded in a computational application (m.knob) under development by FrameNet Brasil. The application, a multilingual pocket tour guide, has three main functionalities: (i) a chatbot that recommends tourist places based on geolocation, modeling of qualia relations, and the interaction with the app; (ii) a Dictiopedia, a multilingual repository of frames and terms of tourism and sports, that is structured in frames, LUs and qualia relations; and (iii) the translator based on frames and qualia structure. As a methodology, first, we work with the constitution of a corpus of real sports sentences, the creation of a human reference translation. Next, the evaluation is performed through the machine translation metrics BLEU, TER, and HTER. These metrics were estimated for the MT system that represents the state of the art (S-Base), the MT system by neural networks with terminological injection in the preprocessing stage (S-Pre) and the MT system by neural networks with terminological injection in the post-editing stage (S-Pós), the last two being developed within the scope of this thesis. The results shows that the terminological injection applied in neural network-based MT systems, through the semantic enrichment with frames and qualia, improves the translation quality of domain specific text, in this case, sports.

Keywords: Frame Semantics. Qualia Structure. FrameNet. Machine Translation. Preprocessing. Post-editing.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURA E SIGLAS

AI	Inteligência Artificial, do inglês, *Artificial Intelligence*
ALPAC	Comitê Consultivo para o Processamento Automático de Língua, do inglês, *Automatic Language Processing Advisory Committee*
Ap.	Apagamento ou Deleção
APE	Pós-edição Automática, do inglês, *Automatic Post-editing*
ARG	Argumento, do inglês, *Argument*
BLEU	Métrica de Avaliação de TM, do inglês, *Bilingual Language Understudy*
CNL	Língua Natural Controlada, do inglês, *Controlled Natural Language*
CNN	Rede Neural Convolucional, do inglês, *Convolutional Neural Network*
CMC	Centro Meteorológico Canadense, do inglês, *Canadian Meteorological Center*
CRF	Campos Aleatórios Condicionais, do inglês, *Conditional Random Fields*
DAISY	Algoritmo de Desamgibuação para a Inferência da Semântica de Y, do inglês, *Disambiguation Algorithm for Inferring the Semantics of Y*
EF	Elemento de Frame, do inglês, *Frame Element* (FE)
EN/En	Inglês, do inglês, *English*
ES/Es	Espanhol, do espanhol, *Español*
FNTi	Algoritmo de Injeção Terminológica da FrameNet, do inglês, *FrameNet Terminology Injection*
F-SEM	Métrica de Avaliação de Tradução por Máquina Baseada na Semântica de Frames
GF	Função Gramatical, do inglês, *Grammatical Function*
GPU	Unidade de Processamento Gráfico, do inglês, Graphics Processing Unit
HTER	Taxa de Edição de Erros de Tradução por Humanos, do inglês, *Human-Targeted Translation Error Rate*
IBM	Corporação Internacional de Máquinas e Negócios, do inglês, *International Business Machines Corporation*
ICSI	Instituto Internacional de Ciência da Computação, do inglês, *International Computer Science Institute*
Ins.	InserçãoMemória de Longo e Curto Prazo, do inglês, *Long Short Term*
LSTM	*Memory*

MEANT	Métrica de Avaliação de Tradução por Máquina
M.KNOB	Base de Dados Multilíngue, do inglês, *Multilingual Knowledge Base*
Mod.	Modificação e Correção
MPos.	Mudança de Posição
MWE	Expressão Multipalavra, do inglês, *Multiword Expression*
NE	Entidade Nomeada, do inglês, *Named Entity*
NLP	Processamento de Língua Natural, do inglês, *Natural Language Processing*
NLU	Compreensão de Língua Natural, do inglês, *Natural Language Understanding*
NP	Sintagma Nominal, do inglês, *Noun Phrase*
OOV	Itens fora do Vocabulário, do inglês, *Out-Of-Vocabulary*
PEMT	Tradução por Máquina com Pós-edição, do inglês, *Post-Editing Machine Translation*
POS	Classe de Palavra, do inglês, *Part of Speech*
PROMT	Projeto de Tradução por Máquina, do inglês, *PROject Machine Translation*
PT	Tipo Sintagmático, do inglês, *Phrase Type*, ou refere-se ao Português
RBMT	Tradução por Máquina Baseada em Regras, do inglês, *Rule-based Machine Translation*
RNN	Rede Neural Recorrente, do inglês, *Recurrent Neural Network*
S-Base	Sistema Base - Sistema de TM convencional com o uso de RNNs
S-Pré	Sistema com Pré-Processamento – Sistema de TM semanticamente enriquecido com frames e qualia, tendo a injeção terminológica no pré-processamento
S-Pós	Sistema com Pós-Edição – Sistema de TM semanticamente enriquecido com frames e qualia, tendo a injeção terminológica na pós-edição
SA	Ativação Propagada, do inglês, *Spreading Activation*
SEH	Hibridização em um Único Sistema, do inglês, *Single-Engine Hybridization*
SL	Língua-fonte, do inglês, *Source Language*
SPE	Pós-edição Estatística, do inglês, *Statistical Post-editing*
STM	Tradução por Máquina Estatística, do inglês, *Statistical Machine Translation*
Sub.	Substituição
TA	Tradução Automática

TAUM	Tradução Automática na Universidade de Montreal, do francês, *Traduction Automatique à l'Université de Montréal*
TL	Língua-alvo, do inglês, *Target Language*
TLG	Teoria do Léxico Gerativo, do inglês, *Generative Lexicon Theory* (GLT)
TM	Tradução por Máquina, do inglês, *Machine Translation* (MT)
TSP	Problema do Caixeiro Viajante, do inglês, *The Traveling Salesman Problem*
UD	Dependências Universais, do inglês, *Universal Dependencies*
UL	Unidade Lexical, do inglês, *Lexical Unit* (LU)
WSD	Desambiguação do Sentido das Palavras, do inglês, *Word Sense Disambiguation*
XML	Linguagem de Marcação Extensível, do inglês, *Extensible Markup Language*
XMEANT	Métrica de Avaliação de Tradução por Máquina

SUMÁRIO

1 INTRODUÇÃO

Atualmente, interações entre computadores e línguas naturais têm se tornado cada vez mais frequentes. Pesquisas na área de Processamento de Língua Natural (*Natural Language Processing* – NLP) são desenvolvidas para conectar diversas áreas da linguagem e as máquinas. Como exemplos desses campos de estudo, temos os Sistemas de Pergunta e Resposta, Sumarização Automática, Tradução Automática, Reconhecimento e Processamento de Fala, Classificação de Documentos, Identificação de Estruturas Discursivas, Reconhecimento de Entidades Nomeadas, Geração de Língua Natural, *Parsing*, Análise de Sentimentos, entre outras. Todos esses estudos impulsionam o desenvolvimento de novas técnicas e tecnologias para aproximar pessoas, melhorar suas vidas e a comunicação através das máquinas.

O presente trabalho insere-se nas áreas de Linguística Computacional, Estudos da Tradução e Linguística Cognitiva. A ideia para a realização desta pesquisa justifica-se a partir de lacunas encontradas em sistemas de Tradução por Máquina (TM)[1], sendo a principal delas a não proposição adequada de equivalentes de tradução para termos de domínio específico, por exemplo os esportes, quando lidamos com palavras polissêmicas.

Os primeiros sistemas de tradução por máquina trabalhavam com um processo de tradução palavra-por-palavra. Detectou-se que questões de reordenação e estruturação sintática se tornavam um problema para eles, além de não considerarem o contexto como parte do processo. Outros sistemas e algoritmos surgem e começam a lidar com o processo de reordenamento sintático e a tradução de frases nas línguas-alvo, a partir de expressões com tradução conhecida. Na última década, a TM passou pela era das ferramentas estatísticas (KOEHN, 2010) e, posteriormente, passou a incorporar Redes Neurais (CHO *et al.*, 2014; SUTSKEVER *et al.*, 2014; BAHDANAU *et al.*, 2015) em seus sistemas de tradução.

Bentivogli *et al.* (2016) apresentam um estudo de TM avaliativo e comparativo entre três abordagens de TM estatísticas baseadas em frases e uma abordagem baseada em redes neurais. A avaliação é feita em cima do par de língua inglês-alemão, dado o fato de essas línguas apresentarem muitos desafios na TM. Os autores utilizam as métricas mTER e HTER de avaliação de TM. Ao analisarem a questão do tamanho das sentenças, Bentivogli *et al.* (2016) observam que no sistema de NMT a qualidade da TM permanece em sentenças maiores, em comparação com os sistemas estatísticos baseados em frases, embora o tamanho grande de

[1] Nesta tese, considera-se que os termos Tradução por Máquina (TM) e Tradução Automática (TA) são sinônimos.

sentenças possa influenciar na piora de qualidade de ambos os tipos de sistemas. Já para os erros morfológicos, lexicais, de reordenamento, e de colocação verbal, os autores apontam que

> As saídas do sistema NMT contém menos erros de morfologia (-19%), menos erros lexicais (-17%) e substancialmente menos erros de ordem de palavras (-50%) do que seu concorrente mais próximo para cada tipo de erro; [...] em relação à ordem das palavras, o NMT mostra uma melhora impressionante na colocação dos verbos (-70% de erros). (BENTIVOGLI *et al.*, 2016, p. 265)[2,3].

Com base no estudo realizado por Bentivogli *et al.* (2016), nota-se que os algoritmos de tradução automática, em especial os que trabalham com redes neurais (NMT), vêm apresentando resultados cada vez melhores no que diz respeito às regras sintáticas, reordenamento de palavras, tamanho das sentenças, aspectos morfológicos e melhores traduções a partir de alta frequência de ocorrência de expressões traduzidas, extraídas de corpora e textos paralelos, além da aplicação de redes neurais.

Koehn e Knowles (2017) apontam alguns desafios que a Tradução Automática por Redes Neurais apresenta. Dentre eles, os autores destacam as peculiaridades de tradução de domínios específicos, sugerindo técnicas de adaptação de domínio como alternativas possíveis. Eles colocam também a questão da quantidade de dados no treinamento das redes neurais, destacando que os sistemas de NMT podem encontrar problemas na tradução de línguas com poucos recursos. Outros desafios recaem sobre as palavras raras ou pouco frequentes, sentenças muito longas, os modelos de alinhamento e as técnicas de Busca em Feixe. Ainda acerca da tradução de domínios específicos, os autores postulam que

> Um desafio conhecido na tradução é que, em diferentes domínios, as palavras têm traduções diferentes e o significado é expresso em estilos diferentes. Portanto, uma etapa crucial no desenvolvimento de sistemas de tradução automática direcionados a um caso de uso específico é a adaptação de domínio. (KOEHN; KNOWLES, 2017, p. 29).[4]

Portanto, partindo do desafio que a TM encontra na tradução de domínios específicos, esta tese se coloca a resolver essa situação ao propor a modelagem semântica (frames e relações qualia) do domínio específico e sua incorporação a modelos de TM em etapas de pré-

[2] "NMT output contains less morphology errors (-19%), less lexical errors (-17%), and substantially less word order errors (-50%) than its closest competitor for each error type; […] concerning word order, NMT shows an impressive improvement in the placement of verbs (-70% errors)." (BENTIVOGLI *et al.*, 2016, p. 265).

[3] Todas as traduções de citações nesta tese são de nossa autoria.

[4] "A known challenge in translation is that in different domains,6 words have different translations and meaning is expressed in different styles. Hence, a crucial step in developing machine translation systems targeted at a specific use case is domain adaptation." (KOEHN; KNOWLES, 2017, p. 29).

processamento e pós-edição, oferecendo um tratamento adequado à semântica das línguas no processo tradutório.

Dois conceitos fundamentais para esta tese e que representam os aspectos semânticos a serem modelados e utilizados por sistemas de TM como estratégias de enriquecimento semântico são frames (FILLMORE, 1982) e relações qualia (PUSTEJOVSKY, 1995).

Entende-se um frame como sendo

> [...] qualquer sistema de conceitos relacionados de tal modo que, para entender qualquer um deles, é preciso entender toda a estrutura na qual se enquadram; quando um dos elementos dessa estrutura é introduzido em um texto, todos os outros elementos serão disponibilizados automaticamente. (FILLMORE, 1982, p. 111).[5]

Os frames podem ser comparados a "cenas" modeladas a partir das nossas experiências com o mundo. Dentro desses frames há participantes (Elementos de Frame – EFs) e características que os distinguem de outros frames. Estabelece-se uma definição para esse frame, havendo um perfilamento de dadas características conforme a particularidade do frame. Posteriormente, dentro desse frame são criadas Unidades Lexicais (ULs), ou seja, palavras dentro da base de dados da framenet associadas a partir de palavras em sentenças (contexto) que evocam um dado frame. A partir do conceito de frame, é possível que se modele uma rede de frames interligados através de relações frame-a-frame. Esses frames ligados compõem uma rede denominada framenet. Há a possibilidade de modelagem de redes de frames de um domínio específico de conhecimento dentro de uma framenet. O conceito de frame e exemplos serão abordados na seção 3.1 Semântica de Frames e suas Aplicações Tecnológicas.

Outro conceito fundamental nesta tese são as relações qualia (PUSTEJOVSKY, 1995). Há quatro tipos de relações qualia que representam aspectos essenciais do significado das palavras. O quale constitutivo é a relação entre um objeto e suas partes constituintes. O quale formal é o que distingue um objeto dentro de um domínio maior. O quale télico se relaciona à função ou propósito do objeto. E por fim, o quale agentivo se coloca nos fatores envolvidos na origem do objeto. Pustejovsky e Jezek (2016) revisitam os conceitos de estrutura qualia propostos por Pustejovsky (1995) e afirmam que

> As qualia codificam aspectos do significado de uma palavra que muitas vezes são atribuídos como conhecimento do mundo pelas teorias linguísticas

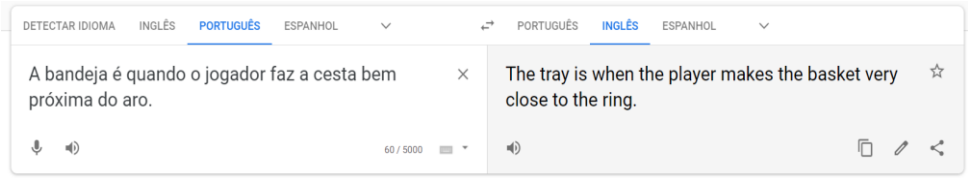
[5] "[...] any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, all of the others are automatically made available". (FILLMORE, 1982, p. 111).

> contemporâneas, isto é, o conhecimento que temos sobre objetos no mundo devido à experiência humana. (PUSTEJOVSKY; JEZEK, 2016, p. 4)[6].

Os autores postulam que, para se identificar o significado de uma palavra, há a necessidade de se conhecer o sistema de representação lexical desta palavra, o que faz com que ela assuma diferentes significados conforme o contexto em que ela se encontre. O conceito de relações qualia será melhor definido e exemplificado na subseção 3.2.2 Níveis de Representação Semântica da TLG.

Observemos na Figura 1 um exemplo de uma sentença de domínio específico, sua tradução automática realizada por um sistema de TM estado da arte (com redes neurais e *transformers* universais) e possíveis problemas com que pretendemos lidar nesta tese.

Figura 1 – Tradução de uma sentença do domínio específico dos Esportes pelo Google Tradutor (NMT – V2)



Fonte: Google Tradutor (https://translate.google.com.br/).

A Figura 1 traz a sentença em língua portuguesa do domínio dos Esportes “A bandeja é quando o jogador faz a cesta bem próxima ao aro.”. Essa sentença é submetida ao Google Tradutor (Sistema de NMT – V2 – com *transformers* universais) que gera a seguinte tradução para a língua inglesa “*The tray is when the player makes the basket very close to the ring.*”. A sentença contém termos específicos dos esportes tais como **bandeja**, **jogador**, **cesta** e **aro**. Partindo disso, há um vocabulário especializado utilizado tanto em português quanto em inglês quando lidamos com o domínio específico dos esportes. Assim, temos um problema com a palavra **bandeja** que, por ser polissêmica e apresentar significados em mais de um contexto, foi traduzida como o objeto retangular achatado utilizado para carregar objetos em cima (***tray***). Entretanto, a tradução de **bandeja** como o tipo de arremesso em que o jogador se aproxima da cesta e usa a tabela para colocar a bola na cesta seria **layup**. Tem-se então um problema de seleção semântica dos equivalentes de tradução adequados ao domínio específico dos esportes.

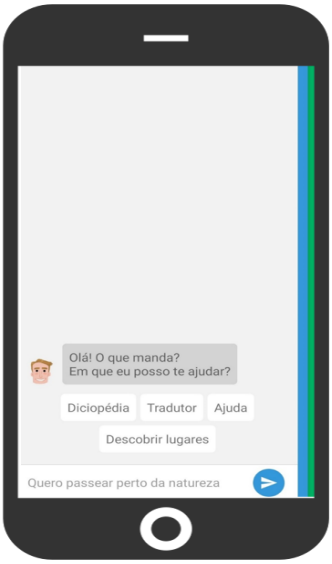
[6] “Qualia encode aspects of a word’s meaning that are often attributed as world knowledge by contemporary linguistic theories, i.e., the knowledge we have about objects in the world due to human experience.” (PUSTEJOVSKY; JEZEK, 2016, p. 4).

Nesta tese, ilustraremos a modelagem do domínio específico dos esportes em termos de frames e relações qualia, além de mostrar como esse tratamento semântico do domínio específico pode contribuir para que um sistema de tradução que utiliza redes neurais ofereça melhores equivalentes de tradução dentro de domínios especializados.

A partir do contexto da TM, dos conceitos apresentados e do problema de tradução no domínio específico dos esportes exemplificado, **o principal objetivo desta tese é propor, a partir da Semântica de Frames (FILLMORE, 1982) e da Teoria do Léxico Gerativo (PUSTEJOVSKY, 1995), uma alternativa de melhoramento semântico da tradução híbrida por máquina para domínios específicos através da modelagem linguístico-computacional do domínio dos Esportes na base de dados da FrameNet Brasil. Para tanto, almeja-se trabalhar na incorporação de frames e relações qualia a dois algoritmos aqui desenvolvidos, sendo ambos enriquecidos semanticamente com frames e relações qualia através de um Sistema de Desambiguação (DAISY) presente em uma das etapas de TM. A diferenciação entre os dois sistemas de TM se coloca no fato de a injeção terminológica ocorrer na etapa de pré-processamento em um deles, e na etapa de pós-edição no outro**.

Considerando o sistema que apresenta melhores resultados, o intuito é que ele seja incorporado como uma função de intérprete pessoal de domínio específico no aplicativo m.knob (*Multilingual Knowledge Base*).

Figura 2 – Interface do Aplicativo m.knob e sua tela de apresentação



Fonte: Aplicativo m.knob. (http://mknob.com).

O m.knob (Figura 2) consiste em uma aplicação computacional cuja função é servir ao usuário como um guia turístico multilíngue de bolso e intérprete pessoal. As três funções principais que integram o aplicativo são: (i) um chatbot que funciona como um sistema de recomendação de atividades e atrações turísticas baseado em frames (PAIVA, 2019), (ii) uma Diciopédia, sendo um repositório multilíngue contendo unidades lexicais, suas definições e relações para oferecer ao usuário conhecimento e informações sobre os domínios do turismo e esportes, cobertos pela aplicação (PERON-CORREA, 2019), além de (iii) um tradutor, foco de aplicação desta tese, também baseado em frames e acrescido de estrutura qualia (PUSTEJOVSKY, 1995). Uma característica interessante a se destacar é o fato de o aplicativo se pautar em domínios específicos, quais sejam o Turismo e os Esportes.

A partir da modelagem linguístico-computacional realizada, **nossa hipótese é a de que melhores equivalentes de tradução são gerados para o domínio específico dos esportes a partir da incorporação da *FrameNet* e das relações qualia ternárias a uma das etapas de um *pipeline* de tradução por máquina**.

Para desempenhar uma avaliação da qualidade da TM em termos da adequação dos equivalentes propostos de domínio específico, estabelecemos uma tradução humana de referência (*gold standard*), feita por um tradutor nativo de língua inglesa, especialista em tradução de domínios específicos, realizamos uma verificação de correspondência semântica entre as sentenças originais em português e a tradução de referência através da inspeção visual de evocação de frames entre o texto original e a tradução de referência. Trouxemos alguns aspectos que contribuem para o desenvolvimento da competência tradutória (HURTADO ALBIR, 2005). Além disso, utilizamos nativos da língua inglesa para atestar a gramaticalidade das traduções de referência. Aplicamos também métricas de avaliação de Tradução por Máquina, sejam elas a BLEU, TER e HTER, que utilizavam a tradução humana de referência (*gold standard*) no processo. A métrica HTER, por sua vez, ainda conta com a revisão das traduções geradas pelo sistema que representa o estado da arte, pelo sistema com injeção em pré-processamento, e pelo sistema com injeção na pós-edição, calculando-se o número de edições necessárias para que essas traduções cheguem o mais próximo da tradução de referência, estabelecendo-se assim, uma métrica comparativa e avaliativa das traduções oferecidas pelos sistemas de TM.

Nesse sentido, nossos objetivos específicos concretizados com a realização desta pesquisa foram:

- Elencar e categorizar Frames, Elementos de Frame (EFs) e Unidades Lexicais (ULs) pertencentes ao domínio dos Esportes Olímpicos, dando ênfase aos nomes de entidade e aos verbos.
- Estruturar e modelar as relações qualia (agentivo, constitutivo, formal e télico) propostas por Pustejovsky (1995) como relações ternárias mediadas por frames no mesmo domínio supracitado, a fim de conseguir melhor refinamento semântico e granularidade do léxico.
- Testar os algoritmos de TM semanticamente melhorados no intuito de checar se há uma mudança significativa na qualidade da tradução para o domínio específico dos esportes, através de métricas de avaliação de TM tais como a BLEU, TER e HTER.

Como resultados alcançados, conclui-se que os sistemas de TM aqui desenvolvidos apresentaram um melhor desempenho na geração de traduções para o domínio específico dos esportes em relação a um sistema estado da arte. O sistema de TM com injeção terminológica na pós-edição foi avaliado com um desempenho na HTER 50% superior ao sistema estado da arte.

Considera-se de fundamental importância, a partir dos estudos aqui propostos, a criação e utilização de métricas de avaliação de tradução que levam em conta aspectos semânticos da tradução como é o caso da F-SEM e da HTER. A F-SEM é uma métrica de avaliação de TM semântica que não requer o uso de traduções de referência ou correções humanas, buscando comparar diretamente o material textual traduzido com o material textual original em termos de frames evocados. Por questões de aplicabilidade (tempo, desenvolvimento computacional, corpora etc), a F-SEM não será utilizada como métrica de avaliação de TM nesta tese, mas almeja-se sua aplicação e uso em trabalhos futuros.

Esta tese está organizada da forma que se expõe a seguir. Apresentaremos, no capítulo 2, o aporte teórico da Tradução por Máquina, suas abordagens, conceitos fundamentais da área, métodos de pré-processamento e pós-edição nos sistemas de TM, além de métricas de avaliação da tradução por máquina. O capítulo 3 traz as teorias linguísticas que fundamentam esta tese, sejam elas a Semântica de Frames e a Teoria do Léxico Gerativo. O capítulo 4 ilustra uma visão geral do aplicativo m.knob (*Multilingual Knowledge Base*), sendo ele o local de futura implementação do sistema de TM enriquecido semanticamente e proposto nesta tese. O capítulo 4 exibe a modelagem realizada na estrutura da base de conhecimento, incluindo a proposição de todos os frames dos esportes e relações semânticas diversas (entre frames, EF-frame, qualia) modelados no âmbito desta tese, além de ontologias utilizadas como bases paralelas de extração de dados e organização computacional de um domínio. O capítulo 5 apresenta a metodologia

experimental com a compilação do corpus de sentenças de teste para os tradutores, a instituição de uma tradução de referência em inglês, o processo de validação dessa tradução de referência e o funcionamento da métrica HTER na avaliação de traduções geradas por sistemas de TM. Por fim, no capítulo 6, são introduzidos o sistema de desambiguação de frames (DAISY) e os sistemas de TM aqui desenvolvidos, sendo um com injeção terminológica no pré-processamento, e o outro com injeção terminológica na pós-edição. Conclui-se o capítulo 6 com avaliação dos sistemas de TM através das métricas BLEU, TER e HTER, passando-se à análise da avaliação de TM dos sistemas aqui estudados. Por fim, apresenta-se a conclusão.

2 PRINCÍPIOS DA TRADUÇÃO POR MÁQUINA

A tradução por máquina (TM), ou tradução automática (TA), configura-se atualmente como uma forma amplamente utilizada para facilitar o acesso a conteúdo em outros idiomas, possibilitando a comunicação entre as pessoas das mais diversas culturas, contribuindo em transações comerciais, aprendizagem de línguas estrangeiras etc. Na área de TM, "tradução pode ser definida basicamente como a tarefa de transformar um texto escrito existente em uma língua-fonte em um texto equivalente em uma língua diferente, a língua-alvo" (GOUTTE *et al.*, 2009, p. 2)[7].

A noção de tradução adotada aqui é a que considera os contextos e aspectos pragmáticos e funcionais desempenhados pelas línguas fonte e alvo no processo tradutório. Harvey (2002, p. 42) postula o conceito de equivalência funcional de tradução em que "esse processo envolve encontrar um referente na língua-alvo que desempenha uma função semelhante. Essa é uma adaptação intercultural." (HARVEY, 2002, p. 42)[8]. Partindo da proposição de equivalência funcional elaborada por Harvey (2002), quando os termos equivalência ou equivalentes de tradução forem utilizados nesta tese, estamos nos referindo à tradução enquanto um processo dialógico não apenas entre línguas, mas entre culturas, considerando a correspondência funcional entre as duas línguas e todo o contexto de produção e recepção envolvido no processo tradutório.

Definir a qualidade de uma tradução é uma tarefa árdua, pois há numerosos fatores que influenciam o fato de se assumir se uma tradução é boa ou ruim. Segundo Hutchins e Somers (1992), os julgamentos são subjetivos e levam em consideração alguns aspectos como a fidelidade, a inteligibilidade, o estilo apropriado, a exatidão, o registro e o público-alvo. Tais características demonstram como a tradução, sendo realizada por humanos ou por máquinas, sofre críticas a todo momento, o que torna a tarefa de se traduzir bem um texto um desafio.

As ferramentas de TM têm buscado oferecer traduções melhores a partir de técnicas variadas, aplicadas através de seus algoritmos. Uma grande quantidade de dados é utilizada a partir de corpora multilíngues compilados, além da utilização de redes neurais no processo tradutório, o que tem contribuído substancialmente para a melhoria dos processos e para a proposição de melhores equivalentes de tradução. Entretanto, Hutchins e Somers (1992) já

---

[7] "Translation is defined as the task of transforming an existing text written in a source language, into an equivalent text in a different language, the target language." (GOUTTE *et al.*, 2009, p. 2).
[8] "Ce procédé consiste à trouver dans la langue d'arrivée un référent qui remplit une fonction similaire. Il s'agit d'une adaptation interculturelle." (HARVEY, 2002, p. 42).

salientavam que os desafios na área de TM não são computacionais em si, mas muito mais linguísticos ao mencionarem

> [...] os problemas de ambiguidade lexical, de complexidade sintática, de diferenças no vocabulário entre línguas, de construções elípticas e "não-gramaticais", de extração do "significado" de sentenças e textos da análise de signos escritos e produzir sentenças e textos em outro conjunto de símbolos linguísticos com um significado equivalente. (HUTCHINS e SOMERS, 1992, p. 2)[9].

Os autores apontaram, já há quase três décadas, alguns dos problemas linguísticos que dificultam o trabalho dos algoritmos de tradução por máquina. Fora os problemas apontados, ainda há questões de terminologia específica, palavras escritas ou digitadas incorretamente, entidades nomeadas, neologismos, entre outros. As redes neurais aplicadas a sistemas de TM passam a contribuir nas tentativas de solução dos problemas linguísticos de tradução apontados. Entretanto, Poibeau (2017) destaca que

> [...] primeiro, treinar redes neurais para a tarefa ainda é difícil devido à sua complexidade, especialmente o número de parâmetros que precisam ser ajustados automaticamente. Isso levou a vários problemas de eficiência. Em segundo lugar, palavras desconhecidas (ou seja, palavras não incluídas nos dados de treinamento) geralmente não são processadas com precisão (ou são simplesmente ignoradas) nesta abordagem. Finalmente, grupos de palavras às vezes não são traduzidos, levando a traduções estranhas e imprecisas. (POIBEAU, 2017, p.126)[10].

Partindo do fato de a TM apresentar problemas linguísticos no desempenho de seus algoritmos, esta tese debruça-se no tratamento da ambiguidade lexical, visto que um dos objetivos do trabalho é buscar soluções de melhorias na tradução por máquina para questões como a ambiguidade lexical para nomes de entidade em domínios terminológicos específicos como os Esportes, por exemplo.

Passemos à seção 2.1 na qual será apresentado um breve histórico da TM.

---

[9] "[…] the problems of lexical ambiguity, of syntactic complexity, of vocabulary differences between languages, of elliptical and 'ungrammatical' constructions, of, in brief, extracting the 'meaning' of sentences and texts from analysis of written signs and producing sentences and texts in another set of linguistic symbols with an equivalent meaning." (HUTCHINS; SOMERS, 1992, p. 2).

[10] "[…] First, training neural networks for the task is still difficult due to their complexity, especially the number of parameters that have to be automatically adjusted. This led to various efficiency problems. Second, unknown words (i.e., words not included in the training data) are generally not processed accurately (or are just ignored) in this approach. Finally, groups of words are sometimes not translated, leading to strange and inaccurate translations." (POIBEAU, 2017, p.126).

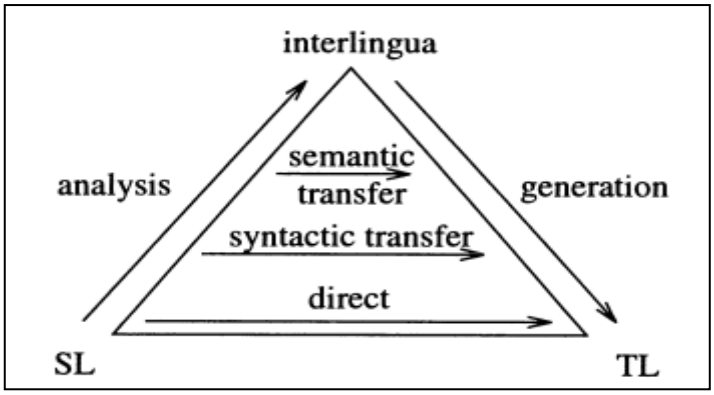
2.1 BREVE HISTÓRICO DA TRADUÇÃO POR MÁQUINA (TM)

Os primeiros trabalhos em tradução por máquina (TM), também denominada "automática" ou "mecânica", datam da década de 1940. Surgem a partir de esforços britânicos de decodificação de mensagens russas durante a Segunda Guerra Mundial. Conforme ilustra Koehn (2010), Waren Weaver foi um dos pioneiros na área de TM em 1947. Ao se deparar com um artigo em russo, afirmou que se tratava de um texto em inglês codificado em símbolos estranhos e que o próximo passo seria decodificá-lo.

Hutchins e Somers (1992) expõem que, em 1954, Leon Dostert colaborou com a IBM (*International Business Machines Corporation*) em um projeto que se tornou a primeira demonstração pública de um sistema de TM. Uma amostra de sentenças em russo foi traduzida para o inglês fazendo uso de um vocabulário de 250 palavras e seis regras gramaticais. Mesmo a demonstração não tendo apresentado relevância científica, significou muito para que novos investimentos e projetos começassem a surgir nos Estados Unidos e na União Soviética. A primeira geração de sistemas de TM (entre as décadas de 1950 e 1970) operava com uma abordagem de tradução direta, ou seja, um texto é dividido em palavras, que são traduzidas e corrigidas superficialmente em relação a sua morfologia e seguindo poucas regras sintáticas. Ainda na década de 1960, apoiadores da TM nos Estados Unidos formam o Comitê Consultivo para o Processamento Automático de Língua (*Automatic Language Processing Advisory Committee* – ALPAC), que, em 1966, lança um dossiê concluindo que a tradução por máquina era lenta, pouco precisa e aparentava ser mais dispendiosa que a tradução realizada por humanos.

Após o Dossiê ALPAC, ocorre a segunda geração de TM (nas décadas de 1970 e 1980), que empregava métodos mais indiretos como a interlíngua ou aqueles baseados em transferência. Nos sistemas baseados em transferência, determina-se inicialmente a estrutura gramatical da sentença, passando-se à manipulação de construções e não apenas de palavras. Isso contribuiu para uma conversão mais refinada na ordem de palavras da tradução. Já os sistemas baseados em interlíngua analisavam os dados como uma metalinguagem e, a partir de regras universais, transformavam os fragmentos de origem em fragmentos traduzidos. Trujillo (1999) traz o triângulo de Vauquois como representação desse sistema de interlíngua e pode ser observado na Figura 3.

Figura 3 – Triângulo de Vauquois e um sistema de TM que considera a interlíngua



Fonte: Translation Engines: Techniques for Machine Translation (TRUJILLO, 1999, p. 6).

A Figura 3 traz uma representação neutra de interlíngua. A partir dessa concepção abstrata, não há a necessidade de uma etapa única de transferência. Trujillo (1999) explica que, na direção vertical do triângulo, está ilustrada a dimensão do esforço necessário para a análise do conteúdo e geração da tradução. Parte-se de uma língua-fonte (*Source language* – SL), à esquerda do triângulo, para uma língua-alvo (*Target language* – TL), à direita do triângulo. Já as linhas horizontais, ilustradas como a base do triângulo e linhas paralelas à base, podem ser compreendidas como a quantidade de esforço necessário para a transferência. Na linha horizontal da base do triângulo, observamos uma tradução direta, mais palavra-por-palavra. Sendo mais próxima da base do triângulo, infere-se que o esforço de transferência palavra-por-palavra ou lexical é menor. Na linha horizontal intermediária, temos demonstrada a transferência sintática, ou seja, são consideradas as regras sintáticas no processo de tradução e transferência e o esforço de transferência já é um pouco maior. Já a linha horizontal mais próxima do ápice do triângulo lida com a transferência semântica, exigindo um alto esforço no processo. O ápice do triângulo pode ser apontado como um estágio abstrato de interlíngua, difícil de ser atingido e que demandaria um esforço muito elevado no processo de transferência. Uma diferença marcante entre uma abordagem de TM baseada em transferência para a abordagem de interlíngua se relaciona ao fato de que, no sistema de interlíngua, é possível se trabalhar com mais de um par de línguas, visto que esse sistema é uma manipulação independente de línguas.

Como apontam Hutchins e Somers (1992), na década de 1970 surge o projeto TAUM (Tradução Automática na Universidade de Montreal[11]) no Canadá que propõe um sistema de tradução por transferência sintática no par inglês-francês. A partir dessa proposta, eles aplicam

[11] Do francês, *Traduction Automatique à l'Université de Montréal*.

tal mecanismo em um sistema de tradução chamado *Météo* voltado à tradução de previsões do tempo. O experimento obtém sucesso devido ao vocabulário restrito e a pouca sintaxe exigida no contexto das previsões do tempo.

Em 1981, Makoto Nagao desenvolve uma abordagem voltada à TM baseada em exemplos (*Example-based*) ou baseada em memória (*Memory-based*). Seu grande benefício é o acesso rápido a grandes bancos de dados de corpora textuais. Esta abordagem envolve o fato de se encontrarem expressões e exemplos análogos, ou seja, como expressões ou frases foram traduzidos anteriormente. Trata-se de processos de extração de palavras e frases equivalentes a partir de textos bilíngues em paralelo produzidos por humanos.

Koehn (2010) expõe que, ainda na década de 1980 e início de 1990, pesquisadores da IBM dão início a aplicações de TM envolvendo grandes quantidades de dados e corpora, caracterizando a abordagem como sendo tradução por máquina estatística. Ela lida com grandes quantidades de textos em corpora paralelos. Quanto mais textos traduzidos, maior será a frequência de que determinado termo possa ser melhor traduzido de determinada forma, o que traz mais qualidade para a tradução e maiores chances de que os equivalentes de tradução sugeridos sejam a melhor opção.

O primeiro modelo da IBM (IBM1) baseava-se em palavras (*Word-based*), isto é, as sentenças traduzidas consideravam as palavras soltas. Esse modelo de TM pode ser chamado de tradução lexical e possui um alinhamento implícito, representando um mapeamento de palavra a palavra entre a língua-fonte e a língua-alvo. Houve outros quatro modelos propostos pela IBM (IBM2, IBM3, IBM4 e IBM5) com tentativas de melhorias em certos aspectos do modelo inicialmente desenvolvido.

Ao perceberem que as palavras não seriam os melhores candidatos como menores unidades para tradução, passou-se a considerar frases (*Phrase-based*). A sentença de entrada era segmentada em frases, que seriam traduzidos considerando seus componentes internos e passando por um processo de alinhamento se necessário. Considerar os segmentos como frases no processo de tradução contribuiria para que certos problemas de ambiguidade fossem resolvidos no processo tradutório, visto que, quanto maior a frequência daquela frase-alvo como sendo equivalente de tradução humana para uma frase-fonte, maiores seriam as chances de aquela frase-alvo ser o equivalente de tradução mais adequado.

Ainda dentro da abordagem estatística voltada à TM, temos os modelos de língua estatísticos baseados em *n-grams*. Sparavigna e Marazzato (2015) apresentam uma definição de *n-gram* como sendo “[...] uma sequência contígua de n-itens de uma sequência fixa de texto ou fala. Os itens podem ser fonemas, sílabas, letras, palavras ou pares-base conforme a

aplicação". (SPARAVIGNA; MARAZZATO, 2015, p. 1)[12]. Trata-se de um modelo de língua que se baseia na distribuição de probabilidades de sequências de palavras. A partir de uma palavra inicial, estima-se a probabilidade de ela ser seguida por uma outra.

Outro tipo de abordagem de TM estatística surge na tentativa de resolver problemas construcionais que a abordagem baseada em frases não conseguiu desempenhar satisfatoriamente. Trata-se de um modelo de TM estatística baseado em sintaxe (s*yntax-based* ou *tree-based*). A ideia desse modelo é explorar, através da sintaxe, como as palavras e as frases se associam no texto-fonte e tentar gerar um texto-alvo equivalente traduzido que consiga abordar tal detalhamento sintático. Acreditava-se ser viável fazer uma mesclagem dessa abordagem baseada em sintaxe com o método baseado em regras, o que solucionaria de uma vez por todas os problemas de alinhamento. Seu maior problema decorreu do fato de não haver *parsers* bons o suficiente que contribuíssem para a rotulagem adequada das categorias sintáticas.

As mais recentes abordagens voltadas à TM têm sido baseadas na utilização de redes neurais, portanto, são denominadas Tradução por Máquina por Redes Neurais (*Neural MT*). As redes neurais surgem na área de Inteligência Artificial e se baseiam nos princípios biológicos do funcionamento das células nervosas. A partir da associação feita às redes neurais animais e aos pesos e forças de conexão entre unidades, exploremos seu conceito e sua utilização na área de TM. Koehn define uma rede neural como "[...] uma técnica de aprendizado de máquina que recebe várias entradas e prevê saídas." (KOEHN, 2017, p. 11)[13]. O funcionamento e os tipos de TM por Redes Neurais serão apresentados e discutidos na seção 2.4 Modelos Neurais de TM. Passemos agora à seção 2.2 com Modelos Heurísticos de TM.

2.2 MODELOS HEURÍSTICOS DE TM

O termo "heurística" vem do grego antigo *εὑρίσκω*, transliterado como *heurísko*, significando "eu acho" ou "eu encontro". A abordagem heurística descreve a capacidade humana de descobrir, criar ou resolver problemas mediante à experiência. Kahneman e Tversky (1974) consideram que

[12] "[…] an n-gram is a contiguous sequence of n-items from a given sequence of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application." (SPARAVIGNA; MARAZZATO, 2015, p. 1).
[13] "A neural network is a machine learning technique that takes a number of inputs and predicts outputs." (KOEHN, 2017, p. 11).

> [...] as pessoas se apoiam em um número limitado de princípios heurísticos que reduzem as tarefas complexas de avaliar probabilidades e prever valores a operações de julgamento mais simples. Em geral, essas heurísticas são bastante úteis, mas às vezes levam a erros graves e sistemáticos. (KAHNEMAN; TVERSKY, 1974, p. 1124)[14].

A partir das considerações mencionadas pelos autores, nota-se que as técnicas heurísticas são estratégias práticas utilizadas na obtenção de resultados rápidos. Entretanto, há uma contraparte negativa apontada, visto que há uma generalização em cima dos resultados, o que pode levar a inúmeros vieses e à padronização de erros. Na ciência da computação, o termo dicionarizado pode ser definido como "prosseguir para uma solução por meio de tentativa e erro ou por regras definidas apenas de maneira vaga" (HOBSON, 2001, p. 205)[15]. Trata-se de estabelecer critérios, métodos ou princípios que são utilizados para se atingir um objetivo, priorizando a velocidade no processamento e a busca rápida por resultados, mesmo que aproximados, deixando de lado alguns preceitos importantes para o processo em si como a qualidade, a integridade, a acurácia e a precisão dos resultados. Um dos exemplos que podem ilustrar o uso de heurísticas na análise é o problema do caixeiro viajante (*The Traveling Salesman Problem* – TSP). Dada uma lista de cidades e as distâncias entre cada par de cidades, qual é a menor rota possível para se visitar cada cidade e retornar à cidade original sem passar novamente por alguma das cidades já visitadas? Partindo desse problema, vejamos a Figura 4 que ilustra a situação.

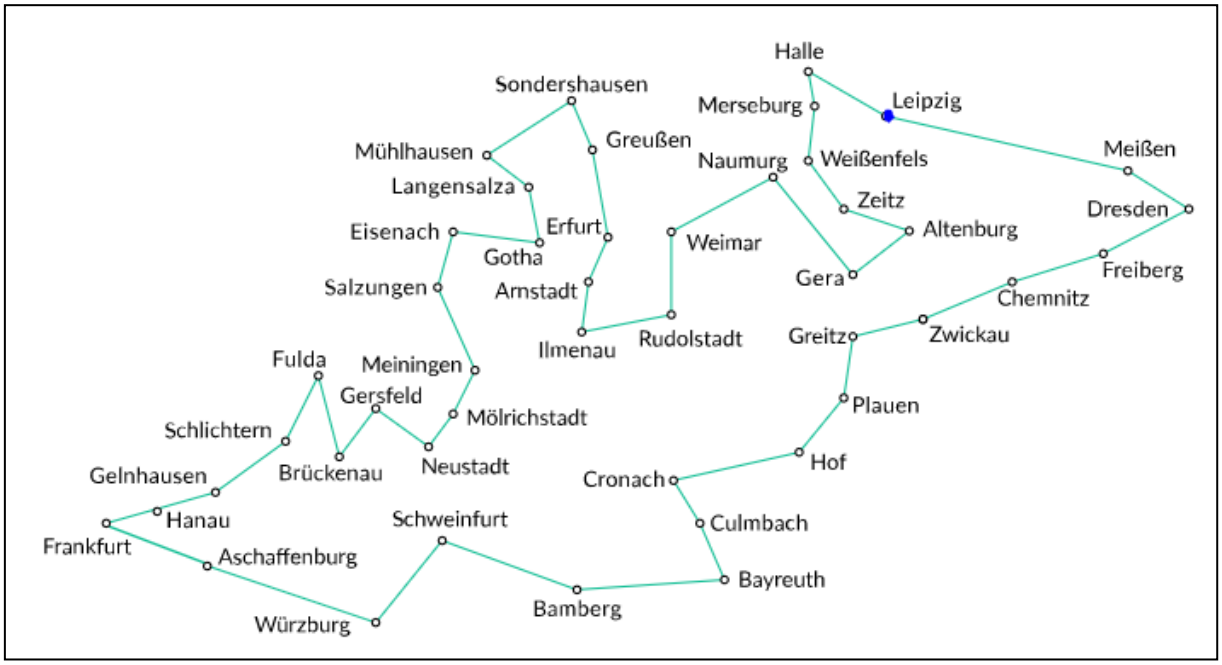
Na Figura 4, temos um diagrama representando 46 cidades da Alemanha. Observamos uma cidade inicial chamada Leipzig marcada por um círculo azul. A partir de Leipzig, traçou-se um trajeto entre as cidades, levando em conta a menor distância entre elas e, a partir disso, houve uma proposição de qual seria o menor trajeto entre as 46 cidades alemãs, até que o ponto final do percurso fosse a cidade de Leipzig novamente. Notemos que nenhuma das cidades foi "visitada" mais de uma vez como solicitado na pergunta-problema. Trata-se de um problema combinatório e, segundo Fox (2019), para que seja resolvido, deveria se aplicar uma "abordagem de força bruta" (*brute force approach*), ou seja, o algoritmo deveria tentar todos os caminhos possíveis entre cidades, medir a distância entre cada caminho e, aí sim, pegar a menor distância possível. A heurística utilizada na resolução do problema passa pela construção

[14] "[…] people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors." (KAHNEMAN; TVERSKY, 1974, p. 1124).

[15] "Comput. Proceeding to a solution by trial and error or by rules that are only loosely defined." (HOBSON, 2001, p. 205).

de um circuito. Parte-se de um princípio de deslocamento para o vizinho mais próximo, até que todas as cidades estejam cobertas pelo viajante. O espaço de procura aumenta exponencialmente à medida em que o número de cidades possíveis também aumenta.

Figura 4 – Diagrama que ilustra o menor trajeto entre 46 cidades alemãs



Fonte: Using Heuristics (FOX, 2019).

Na Tradução por Máquina, há abordagens heurísticas que são baseadas em regras (*Rule-based*). Essas abordagens tomam as informações linguísticas da entrada e da saída com base em dicionários e gramáticas. Duas abordagens que se encaixam no modelo de TM baseado em regras são: a Tradução Direta (*Direct Translation*) e a Tradução baseada em transferência (*Transfer-based*). Na Tradução Direta, concebe-se a tradução como sendo palavra por palavra, com a utilização de dicionários bilíngues e podendo haver regras simples de reordenamento. Já o modelo baseado em transferência é apresentado por Jurafsky e Martin (2008) como um modelo que

> [...] pressupõe um *parsing* do idioma de origem e é seguido por uma fase de geração para realmente criar a sentença de saída. Assim, nesse modelo, a TM envolve três fases: análise, transferência e geração, em que a transferência preenche a lacuna entre a saída do *parser* do idioma de origem e a entrada para o gerador do idioma de destino.[16] (JURAFSKY; MARTIN, 2008, p. 884).

[16] "The transfer model presupposes a parse of the source language, and is followed by a generation phase to actually create the output sentence. Thus, on this model, MT involves three phases: analysis, transfer, and generation, where transfer bridges the gap between the output of the source language parser and the input to the target language generator." (JURAFSKY; MARTIN, 2008, p. 884).

O modelo apresentado envolve a necessidade de um conjunto de regras para a transferência lexical e sintática. O *parsing* é uma etapa do processamento linguístico que pode ser definida como a ação de se tomar uma entrada e atribuir a ela alguma estrutura linguística. O termo pode ser compreendido como uma análise morfológica, sintática ou semântica que atribui uma rotulação linguístico-estrutural a um determinado segmento. Passemos então a exemplos de aplicações de TM que lidam com métodos heurísticos em seu processamento.

Em 1975, o grupo TAUM, com apoio do governo canadense, começou a desenvolver na universidade de Montreal um projeto denominado *Météo*, voltado para a tradução automática de previsões do tempo do inglês para o francês. A iniciativa partiu do interesse em se disponibilizar para a população as previsões do tempo referentes aos dados climáticos de todo o país em língua francesa. Era a primeira vez que um produto resultante de tradução automática seria disponibilizado ao público.

As informações climáticas compiladas pelo Centro Meteorológico Canadense (CMC) eram traduzidas para o francês e repassadas à população via estações de rádio e jornais. Na época, o sistema conseguia traduzir "previsões regionais, previsões marítimas e previsões destinadas especificamente a agricultores e velejadores." (THOUIN, 1981, p. 40)[17]. Acerca da caracterização dos dados linguísticos do *Météo*, Hutchins e Somers (1992) expõem que

> Os dados linguísticos do *Météo* consistem em três dicionários bilíngues de 'expressões idiomáticas', nomes de lugares e vocabulário geral (meteorológico) e três módulos de processamento para a análise sintática do inglês, a geração sintática do francês e a geração morfológica do francês. (HUTCHINS; SOMERS, 1992, p. 209)[18].

O grupo TAUM utilizava uma abordagem de transferência para suas pesquisas em geral sobre TM. Entretanto, para o *Météo*, o design de tradução "direta" também foi amplamente utilizado, visto que tanto o inglês quanto o francês compartilhavam de relatórios meteorológicos em estilo telegráfico com um vocabulário restrito, eliminação de certas partículas devido ao gênero textual em questão e pouca variação morfológica. As traduções eram feitas em cima de 80% do material linguístico, visto que 20% das entradas apresentavam "[...] palavras com erros ortográficos, caracteres borrados na transmissão, palavras que não estavam no dicionário, inglês

[17] "[...] regional forecasts, maritime forecasts and forecasts aimed specifically at farmers and boaters." (THOUIN, 1981, p. 40).

[18] "The linguistic data of *Météo* consist of three bilingual dictionaries for 'idioms', place names and general (meteorological) vocabulary, and three processing modules for the syntactic analysis of English, the syntactic generation of French, and the morphological generation of French." (HUTCHINS; SOMERS, 1992, p. 209).

com problemas, estruturas sintáticas desconhecidas para o *parser* etc." (CHANDIOUX, 1976, p. 28)[19]. As mensagens que não eram possíveis de serem codificadas pelo terminal eram encaminhadas para um tradutor humano para que pudesse analisar o conteúdo.

O processo de tradução utilizado pelo *Météo* passava por um Sistema-Q (*Q-System*), que trabalhava com um conjunto restrito de regras. Chandioux (1976) indica que o sistema possuía quatro fases, sejam elas, a fase do dicionário de expressões idiomáticas, a fase do dicionário principal, a fase do *parser* e a fase do gerador. O autor expõe que o dicionário de expressões idiomáticas continha 300 entradas (ex.: *clear period* ➔ *éclaircie*) e lidava com termos específicos do clima e com nomes de locais que possuíam traduções próprias (ex.: *Lake St. Claire* ➔ *lac Ste Claire*). Nessa fase, havia a expansão de abreviações, identificação de expressões idiomáticas referentes ao clima e o processamento de nomes de lugares. O dicionário principal continha 1200 entradas e servia de base para o funcionamento das fases de *parsing* e geração da tradução. Nessa etapa, eram atribuídas a cada palavra possíveis categorias sintáticas, e para cada categoria, uma tradução possível. A fase do *parser* lidava com a análise sintática de sentenças curtas e estruturalmente simples. Durante o *parsing*, eram atribuídos rótulos de identificação aos segmentos de modo a facilitar a conversão para a outra língua. Dentre esses rótulos, os fragmentos eram marcados como contendo datas, horas, temperaturas, e até características linguístico-temporais tais como o aspecto durativo ou pontual da expressão (ex.: *in the morning* ➔ *dans la matinée / this morning* ➔ *ce matin*). A fase de *parsing* contava com aproximadamente 300 regras de reescrita, sendo uma fase sensível ao contexto. A fase geracional já tinha como tarefa decompor as estruturas geradas pelo *parser*, levando em consideração a ordem de palavras e concordância do francês (ex.: *gusty westerly winds* ➔ *vents d'ouest soufflant em rafales*). Nessa etapa, o sistema realizava uma separação dos sintagmas nominais e contava com 300 regras de reescrita, atendendo a aspectos morfológicos (concordância dos adjetivos, elisão, contrações etc.) e estilísticos.[20]

Em 1984, Makoto Nagao cria a uma abordagem de TM baseada em exemplos. Para tal criação, o autor se inspira em um modelo de aprendizagem de línguas estrangeiras em que os alunos memorizam um conjunto de palavras e sentenças em inglês e em japonês. Com base nessa memorização, o aluno pode fazer inferências sobre a estrutura das sentenças da língua a ser aprendida a partir dos exemplos que ele possui. Para a TM, o processo proposto é de oferecer inicialmente uma sentença-exemplo simples em inglês e sua correspondência em japonês. O

[19] "[…] misspelled words, characters blurred in transmission, words not in dictionary, poor English, syntactic structures unknown to the parser, etc." (CHANDIOUX, 1976, p. 28).
[20] Os exemplos de tradução inglês-francês ilustrados aqui foram elencados por Chandioux (1976).

passo seguinte é disponibilizar ao sistema uma segunda sentença-exemplo em inglês com sua equivalente em japonês, havendo a alteração de apenas uma palavra. As substituições de palavras na sentença são realizadas uma por vez e nas posições de sujeito, objeto e complemento. Após a operação de substituição, um humano avalia a aceitabilidade ou não da equivalência de tradução representada. Segundo o autor, dessa forma, o sistema obtém "[...] certos fatos acerca da estrutura da sentença e uma correspondência entre palavras do inglês e do japonês." (NAGAO, 1984, p. 174)[21]. As substituições não eram feitas com os verbos, visto que cada verbo pode apresentar características específicas em relação à estrutura sentencial, ou seja, um padrão de valência específico e diferenciado a cada uso. Havia a utilização de dicionários e tesauros. Um tesauro é um dicionário de ideias afins e relações entre palavras, havendo sinônimos, antônimos, conceitos de hiperonímia e hiponímia, além de relações parte-todo, entre outras. Com uma determinada sentença a ser traduzida, o sistema propõe uma sugestão de tradução baseada nas inferências e no modelo. Trabalha-se com tesauros em uma etapa em que o sistema verifica a substituibilidade de um termo por outro oferecido, dada sua relação estabelecida com outros termos no tesauro. Portanto, essa abordagem se torna muito dependente de uma grande quantidade de informações modeladas no sistema via exemplos análogos traduzidos, além de uma grande vinculação aos tesauros. Passemos agora à seção 2.3 com Modelos Estatísticos de TM.

2.3 MODELOS ESTATÍSTICOS DE TM

A tradução por máquina estatística baseia-se na compilação de grandes quantidades de corpora paralelos entre a língua-fonte e a língua-alvo e, com a frequência de ocorrências de um determinado termo ou frase, o sistema é capaz de propor um equivalente de tradução mais adequado, dado o contexto. Nos sistemas de TM, há tentativas de criação de modelos de língua. Modelo de língua pode ser concebido como um modelo matemático probabilístico de como funciona uma determinada língua. Os modelos de língua fazem a utilização de *n-grams*. Kok e Brower (2011) apontam que um unigrama (*unigram*) pode ser analogamente compreendido como uma janela na qual olhamos apenas uma palavra do texto. Já um bigrama (*bigram*) seria a janela conseguindo olhar duas palavras por vez no texto. Partindo desse princípio, um modelo de língua utilizaria *n-grams*, ou seja, combinações possíveis de n palavras e estruturas da língua,

[21] "Certain facts about the structure of a sentence and correspondence between English and Japanese words." (NAGAO, 1984, p. 174).

em que, com a utilização de determinada palavra, pode-se prever quais outras palavras possíveis podem coocorrer em um texto. Portanto, seria possível conhecer o funcionamento da língua através da utilização desse modelo.

A TM estatística surge no final da década de 1980 e início da década de 1990 nos laboratórios da IBM através do projeto Candide. Dellapietra e Dellapietra (1994) descrevem os dois objetivos centrais de tal projeto. Primeiro, buscavam desenvolver um sistema totalmente automático de tradução com amplo vocabulário do francês para o inglês. Seu segundo objetivo era elaborar uma estação de trabalho de tradução interativa que melhorasse a velocidade e produtividade do tradutor humano. Seu sistema possuía um componente de transferência que, incorporando um modelo de língua, estimava a probabilidade de uma estrutura do inglês intermediária. O tamanho do vocabulário do inglês era de 70.000 palavras e do francês 280.000 palavras. Quanto ao aspecto linguístico, havia uma análise de tabelas morfológicas, um tratamento de expressões numéricas e nomes próprios, além de melhorias nas transformações sintáticas e em um modelo bilíngue estatístico de desambiguação de sentidos.

O primeiro modelo de tradução estatística da IBM, o IBM1, considerava palavras isoladas, podendo também ser chamado de tradução lexical. Nesse modelo, havia a necessidade de um dicionário bilíngue que mapeasse os equivalentes de uma língua na outra. Estimava-se uma distribuição de probabilidades da tradução lexical em que uma palavra possuía determinados equivalentes de tradução possíveis a partir da frequência de ocorrência em corpus. Essa distribuição pode ser representada pela Equação 1.

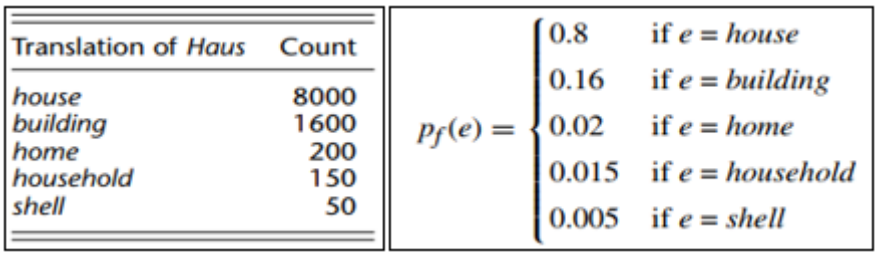
Equação 1 - Fórmula da Distribuição de Probabilidades

$$p_f : e \rightarrow p_f(e)$$

Fonte: Statistical Machine Translation (KOEHN, 2010, p.82).

A Equação 1 indica que uma dada palavra **f** em uma língua-fonte retorna uma probabilidade para cada escolha de tradução **e** em uma língua-alvo, indicando a probabilidade de ocorrência desse equivalente de tradução. Ilustremos essa fórmula através da Figura 5.

Figura 5 – Representação da distribuição de probabilidades para os equivalentes de *Haus* (casa em alemão) em inglês



Translation of *Haus*	Count
house	8000
building	1600
home	200
household	150
shell	50

$$p_f(e) = \begin{cases} 0.8 & \text{if } e = \textit{house} \\ 0.16 & \text{if } e = \textit{building} \\ 0.02 & \text{if } e = \textit{home} \\ 0.015 & \text{if } e = \textit{household} \\ 0.005 & \text{if } e = \textit{shell} \end{cases}$$

Fonte: Statistical Machine Translation (KOEHN, 2010, p.83).

Na Figura 5, à esquerda, temos uma distribuição hipotética da quantidade de ocorrências de tradução para a palavra *haus* do alemão em inglês em corpora paralelos. Nota-se que a sugestão de tradução mais frequente seria *house* em inglês, com 8000 ocorrências, não descartando outras possíveis traduções, ranqueadas com menor frequência, a partir da análise em corpora. À direta na Figura 5, partindo do fato hipotético de se haver 10.000 ocorrências da palavra *haus* (em alemão) no corpus, há uma normalização dos valores de ocorrências em inglês. Seguindo a frequência em corpus, o melhor equivalente de tradução para *haus* seria *house*, com 80% de ocorrências de *house* no corpus paralelo, ficando *building* como segundo equivalente de tradução mais frequente, com 16% de ocorrências, *home* em terceiro lugar com 2% de ocorrências, *household* em quarto lugar com 1,5% de ocorrências, e *shell* em último lugar de possíveis traduções para *haus*, com 0,5% de ocorrências. Considerar a incorporação de processos estatísticos, modelos de língua e corpora nos processos tradutórios traz mais confiabilidade de que os melhores equivalentes de tradução seriam X, Y ou Z, dadas as suas ocorrências reais em corpora com grandes quantidades de dados.

A partir dessa análise estatística de ocorrências dos equivalentes entre línguas nos corpora paralelos, começa-se a estabelecer tabelas de tradução com a distribuição de probabilidades de frequência de ocorrências de cada palavra da sentença em corpora paralelos. Observemos na Tabela 1 essa distribuição.

Tabela 1 – Distribuição de probabilidades para os equivalentes de tradução de quatro palavras

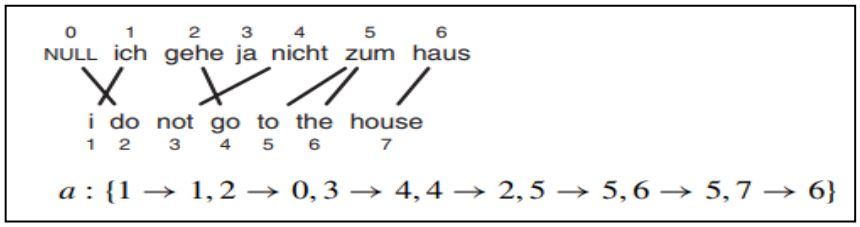
das		**Haus**		**ist**		**klein**	
e	*t*(*e*| *f*)	*e*	*t*(*e*| *f*)	*e*	*t*(*e*| *f*)	*e*	*t*(*e*| *f*)
the	0.7	*house*	0.8	*is*	0.8	*small*	0.4
that	0.15	*building*	0.16	*'s*	0.16	*little*	0.4
which	0.075	*home*	0.02	*exists*	0.02	*short*	0.1
who	0.05	*household*	0.015	*has*	0.015	*minor*	0.06
this	0.025	*shell*	0.005	*are*	0.005	*petty*	0.04

Fonte: Statistical Machine Translation (KOEHN, 2010, p.84).

Na Tabela 1, temos que o equivalente de maior frequência para *das* (alemão) em inglês seria *the*, representando 70% de ocorrências em corpus como equivalente mais comum em inglês. *House* aparece como a palavra mais frequente como tradução de *haus* (alemão), com 80% de frequência. *Is* apresenta 80% de frequência como tradução em inglês para *ist* (alemão). Por fim, *small* e *little* aparecem em corpus como os termos mais frequentes em inglês, com 40% das ocorrências, como a tradução mais frequente para *klein* (alemão). A partir dessas possibilidades de ocorrência, a tradução em inglês mais adequada para a sentença alemã *das Haus ist klein* seria *The house is small/little*[22]. Tal tradução estaria amparada pela frequência de ocorrências dos termos em um contexto oferecido pelos corpora.

A partir desse modelo lexical, começa-se a inserir uma função de alinhamento no algoritmo de tradução, estabelecendo que determinada palavra de uma língua-fonte que ocupa uma posição na sentença vinculada a um número-índice se ligaria a uma outra palavra na língua-alvo que ocupa a mesma posição ou outra diversa, também apresentando um número-índice. Observemos na Figura 6 como funcionaria essa proposição do modelo de alinhamento entre sentenças com índices e ligação entre eles a partir de uma sentença em inglês e alemão que pode ter como uma correspondência em português a sentença “Eu não vou para casa”.

Figura 6 – Representação do modelo de alinhamento



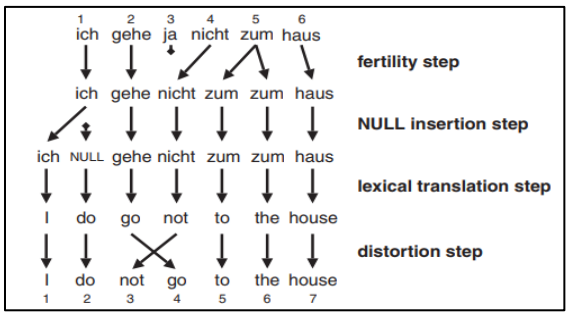
Fonte: Statistical Machine Translation (KOEHN, 2010, p.85).

[22] Traduzida para o português como “A casa é pequena” (Tradução nossa).

Ao analisarmos o modelo de alinhamento proposto entre a sentença em alemão e a sentença em inglês proposto na Figura 6, temos a representação através da função *a: j* ➔ *i*, em que o alinhamento é determinado com uma ou mais palavras do inglês na posição **i** ligadas a uma ou mais palavras do alemão na posição **j**. Não havendo nenhuma equivalência de determinado termo ou estrutura, insere-se no modelo o *token* nulo, para que se efetue a ligação entre estruturas não equivalentes entre línguas. Portanto, estabelece-se que *I* (inglês) seria a tradução de *ich* (alemão), ambos sendo ligados ao índice 1, *do* (inglês) correspondendo ao *token* nulo, *go* (inglês) na posição 4 como equivalente de *gehe* (alemão) na posição 2, *not* (inglês) na posição 3 ligado a *nicht* (alemão) na posição 4, *to* (inglês) na posição 5 e *the* (inglês) na posição 6 ligados à *zum* (alemão) na posição 5, e finalmente, *house* (inglês), na posição 7, ligado a *haus* (alemão), na posição 6. Demonstramos então a proposta do modelo IBM1 com a tradução lexical, o estabelecimento de um modelo de língua e um modelo de alinhamento no algoritmo de tradução. Nesse modelo, os reordenamentos possíveis entre estruturas são todos analisados igualmente.

O IBM2 surge acrescentando o modelo de alinhamento absoluto. Os alinhamentos funcionam como uma forma de lidar com o reordenamento de palavras nas sentenças. Há então a etapa lexical de tradução estatística, acrescida pela etapa seguinte de atribuição de índices/posições para as palavras e havendo o reordenamento. Há uma técnica heurística envolvida nesse processo denominada busca em feixe (*beam search*), ou seja, estabelece-se um parâmetro do tamanho do feixe que é considerado na busca, estipulando uma certa largura possível de hipóteses para as palavras vizinhas, e assim, limitando o número de traduções parciais mantidas por palavra de entrada. O modelo IBM3 passa a incorporar o modelo de fertilidade, no qual as palavras de entrada produzem um número específico de palavras na saída, podendo ser 0, 1 ou mais palavras. Vejamos na Figura 7 as etapas propostas pelo IBM3.

Figura 7 – Representação das etapas propostas pelo modelo IBM3



Fonte: Statistical Machine Translation (KOEHN, 2010, p.101).

Na Figura 7, temos ilustrados os passos propostos pelo IBM3 e o modelo de fertilidade criado. Na proposta do modelo de fertilidade (*fertility step)*, atribui-se a cada palavra um possível equivalente na outra língua como uma fase intermediária do processo. A fertilidade das palavras pode ser compreendida como o número de palavras de saída que são geradas a partir de cada palavra de entrada. Nota-se esse processo de fertilidade na duplicação da palavra alemã *zum* (para a), visto que em inglês ela geraria *to the* (para a) como equivalente. Em seguida, há a fase da inserção de um *token* nulo (*null insertion step*) para ligar possíveis estruturas que não possuem um equivalente direto daquela estrutura de sentença na outra língua. Passa-se à fase de tradução lexical (*lexical translation step*) em que todos os termos da estrutura intermediária são traduzidos seguindo as distribuições de probabilidades e as ocorrências dos *n-grams*. Por fim, há uma fase de distorção (*distortion step*), em que se produz a mesma tradução e alinhamento, mas o sistema realiza uma reordenação, se necessário, a partir da previsão das posições das palavras de saída baseadas nas posições das palavras de entrada.

Ao se testar os modelos e algoritmos com sentenças de entrada mais extensas, detectou-se um problema na distorção oferecida pelo IBM3. A partir daí, o IBM4 emerge com um modelo de alinhamento relativo. Em tal modelo, realiza-se um alinhamento prévio, em que cada palavra é dependente e alinhada a outra palavra anteriormente em núcleos que consideram as palavras próximas, havendo também uma relação estabelecida com as classes de palavra das palavras adjacentes. As palavras que aparecem adjacentes na entrada tendem a continuar próximas na tradução. Há línguas em que o substantivo e adjetivo possuem ordem fixa em determinadas construções, o que faz com que um realinhamento seja necessário.

O IBM5 surge com o propósito de corrigir deficiências apresentadas em modelos anteriores. Algumas dessas deficiências dizem respeito a traduções impossíveis que apresentam probabilidades positivas ou a múltiplas palavras de saída que podem ocupar a mesma posição. O modelo IBM5 prevê que podemos inserir palavras apenas em posições de palavras vagas (*vacant word positions*). A partir dessa proposta, rastreia-se todas essas posições possíveis e apenas permite-se a colocação de palavras em determinadas posições, eliminando assim a questão das deficiências propostas anteriormente.

Baseando-se no princípio de que as palavras na língua-fonte podem ser traduzidas por mais de uma palavra na língua-alvo, representando um problema para determinados sistemas, surgem os modelos baseados em frases. Estes passam a considerar os frases ou unidades multipalavras no processo tradutório. Koehn (2010) sugere algumas vantagens de um modelo de tradução baseado em frases.

> Primeiro, as palavras podem não ser as melhores unidades atômicas para tradução, devido aos frequentes mapeamentos uma-para-muitas (e vice-versa). Em segundo lugar, traduzir grupos de palavras em vez de palavras únicas ajuda a resolver ambiguidades de tradução. Há um terceiro benefício: se tivermos grandes corpora de treinamento, podemos aprender frases úteis cada vez mais longas, às vezes até memorizar a tradução de sentenças inteiras. Finalmente, o modelo é conceitualmente muito mais simples. (KOEHN, 2010, p. 128)[23].

Koehn traz algumas vantagens acerca da utilização de frases no processo da tradução. Semelhante ao que se fazia na tradução lexical, traçam-se tabelas de frequência de tradução com base nas ocorrências em corpora para os frases e cria-se um modelo de reordenamento baseado em distância. Movimentos a longas distâncias na sentença tendem a ser mais dispendiosos no processo de tradução do que aqueles realizados a curtas distâncias.

Um problema que surge na abordagem que toma as frases no processo tradutório são algumas frases que, por serem mais infrequentes, encontram problemas na tradução. Uma tentativa de lidar com essas frases é através da utilização de pesos lexicais atribuídos às palavras que compõem a frase, dando confiabilidade de que aquela frase existe na língua. Outros elementos inseridos nos modelos de tradução são a penalidade de palavra e penalidade de frase. Tais penalidades oferecem ao sistema um fator que penaliza as saídas de tradução que sejam muito curtas ou muito longas no âmbito da palavra e da frase, respectivamente. Esse sistema de penalidades pressupõe que a tradução de saída não deva ser tão mais curta ou mais longa que a entrada, a depender do par de línguas levado em conta no processo tradutório. O reordenamento de palavras passa a ser limitado, considerando-se características especiais do par de línguas. Por exemplo, na tradução do par árabe-inglês, o reordenamento entre sujeito-verbo e adjetivo-nome tem que ser levado em conta, visto que as traduções oferecidas devem ser condizentes com a estrutura das línguas traduzidas.

Ao se depararem com certos problemas, repensaram que a melhor unidade de tradução não seria mais a frase, passando então a levar em conta a sintaxe da sentença como um todo, surgindo assim, modelos de TM estatística baseados em sintaxe. O objetivo dessas técnicas de TM estatística baseada em sintaxe era incorporar representações explícitas da sintaxe nos sistemas de tradução. Começaram então a utilizar unidades sintáticas de tradução, ou seja,

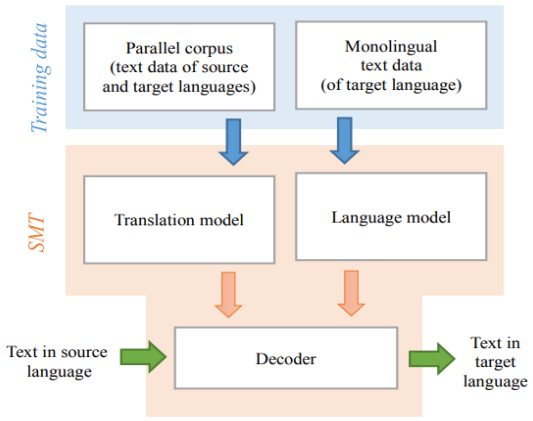
[23] "For one, words may not be the best atomic units for translation, due to frequent one-to-many mappings (and vice versa). Secondly, translating word groups instead of single words helps to resolve translation ambiguities. There is a third benefit: if we have large training corpora, we can learn longer and longer useful phrases, sometimes even memorize the translation of entire sentences. Finally, the model is conceptually much simpler." (KOEHN, 2010, p. 128).

passaram a inserir métodos de *parsing*, analisando as funções sintáticas atribuídas às palavras das sentenças.

Essa abordagem de TM estatística tornou-se amplamente difundida e utilizada por aplicações diversas. O Moses, por exemplo, é um algoritmo de TM estatística que possibilita o treinamento automático de modelos de tradução para pares de língua diversos. Uma coletânea de textos traduzidos se faz necessária, ou seja, uma compilação de corpora paralelos. Uma vez treinado o modelo, o algoritmo de busca tenta encontrar a tradução com maior probabilidade de ocorrência em um número elevado de opções. O Google Tradutor, um dos maiores sistemas de tradução por máquina existentes, passou a incorporar a tradução por máquina estatística a seus modelos a partir de 2006. Outras aplicações de tradução que também passaram a utilizar modelos estatísticos foram o tradutor da Microsoft e o Yandex.

O tradutor Yandex estatístico foi implementado em 2011 pela empresa russa de mesmo nome. Seu sistema de tradução é composto por três módulos: um módulo de tradução, um modelo de língua e um decodificador que podem ser ilustrados na Figura 8.

Figura 8 – Modelo de Tradução por Máquina estatística em três módulos



Fonte: Web-based Automatic Translation: the Yandex Translate API (VAN HEES *et al.*, 2015).

Na Figura 8, temos ilustrados inicialmente no topo os dados de treinamento (*training data*). Os dados em corpora paralelos (*parallel corpus*) da língua-fonte e da língua-alvo servirão de base para o modelo de tradução. Já os dados de treinamento que partem de textos monolíngues da língua-alvo (*monolingual text data*) são necessários para a constituição do modelo de língua (*language model*). O modelo de tradução é simplesmente uma lista de todas as palavras conhecidas na língua-alvo e suas traduções em outra(s) língua(s). O sistema observa as frases correspondentes, os organiza em grupos de palavras e calcula as probabilidades baseadas em ocorrências prévias de tradução. O modelo de língua fica responsável por fornecer

o conhecimento acerca da língua-alvo. Ele garante que a tradução gerada seja compatível com as estruturas utilizadas na língua. Por fim, o decodificador (*decoder*) faz a tradução de verdade. Ele combina a entrada do usuário e o modelo de tradução para gerar as traduções, ranqueadas de acordo com a probabilidade de ocorrência. Vimos então o funcionamento de um sistema estatístico de TM. Passemos à seção 2.4 com os Modelos Neurais aplicados à TM.

2.4 MODELOS NEURAIS DE TM

As Redes Neurais Artificiais (*Artificial Neural Networks*), também conhecidas como sistemas conexionistas, são sistemas computacionais inspirados no sistema nervoso central dos animais, especificamente o cérebro, sendo capazes de aprender tarefas realizadas por humanos tais como classificação, previsão, tomada de decisões, visualização, reconhecimento de padrões, entre outras. Elas adquirem essa capacidade de aprendizado de máquina, testando, errando e corrigindo o erro, a partir de exemplos e da experiência. Aplicações que utilizam redes neurais têm emergido em diversas áreas como a Visão Computacional, Filtros em Redes Sociais, Jogos de tabuleiro e videogames, diagnósticos médicos, entre outras, incluindo questões de Processamento de Língua Natural (NLP) tais como a Tradução por Máquina e o Reconhecimento de Fala.

Como definição de rede neural, Gurney (1997) afirma que

> [...] é um conjunto interconectado de elementos, unidades ou nós de processamento simples, cuja funcionalidade é vagamente baseada no neurônio animal. A capacidade de processamento da rede é armazenada nos pontos fortes ou pesos de conexão entre unidades, obtidos por um processo de adaptação ou aprendizado de um conjunto de padrões de treinamento. (GURNEY, 1997, p. 13).[24]

A partir da definição proposta por Gurney (1997), percebe-se a analogia feita à rede de neurônios animal e à considerável capacidade de aprendizado das redes neurais a partir da experiência e do treinamento. Complementemos essa definição com as colocações de Haykin (2008), ao postular que

> [...] uma rede neural é um processador paralelo massivamente distribuído, composto de unidades simples de processamento que tem uma propensão natural para armazenar conhecimento experimental e disponibilizá-lo para uso. Assemelha-se ao cérebro em dois aspectos: (1) o conhecimento é

[24] "A neural network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the interunit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns." (GURNEY, 1997, p. 13).

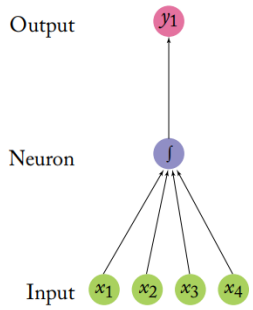
> adquirido pela rede a partir de seu ambiente através de um processo de aprendizado; (2) Os pontos fortes da conexão entre neurônios, conhecidos como pesos sinápticos, são usados para armazenar o conhecimento adquirido. (HAYKIN, 2008, p. 2).[25]

Nota-se com as definições propostas que uma rede neural é um sistema ou processador distribuído que tenta simular as sinapses ou conexões entre suas unidades chamadas de neurônios. A rede é treinada, armazena conhecimento e aprende através de testes, erros e a diferença entre esses últimos e os acertos que gera. O sistema faz com que uma saída se torne uma nova entrada, refazendo análises e, consequentemente, aprendendo com a experiência. Na seção subsequente, apresentamos os conceitos fundamentais envolvidos nos métodos que utilizam redes neurais.

2.4.1 Conceitos Fundamentais

O primeiro dos conceitos relevantes para redes neurais é o de neurônio, representado na Figura 9.

Figura 9 – Um único neurônio com quatro entradas



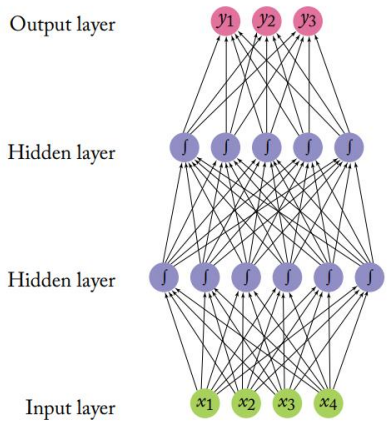
Fonte: Neural Network Methods for Natural Language Processing (GOLDBERG, 2017, p. 41).

A Figura 9 ilustra um neurônio. Há quatro entradas (*Input*) nesse neurônio, representadas pelos círculos verdes nomeados por **x** e um número. O neurônio (*Neuron*) aparece no círculo roxo denominado **∫.** Já a saída (*Output*) é ilustrada pelo círculo rosa chamado **y** seguido de um número. O neurônio processa as entradas que aparecem submetidas ao sistema

[25] "A neural network is a massively parallel distributed processor made up of simple processing units that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: (1) Knowledge is acquired by the network from its environment through a learning process; (2) Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge." (HAYKIN, 2008, p. 2).

e gera uma saída. Vejamos na Figura 10 um dos diversos modelos de redes neurais para entendermos o papel do neurônio, o funcionamento de uma rede neural e os mecanismos subjacentes às redes neurais.

Figura 10 – Rede Neural *Feed-forward* com duas camadas ocultas



Fonte: Neural Network Methods for Natural Language Processing (GOLDBERG, 2017, p. 42).

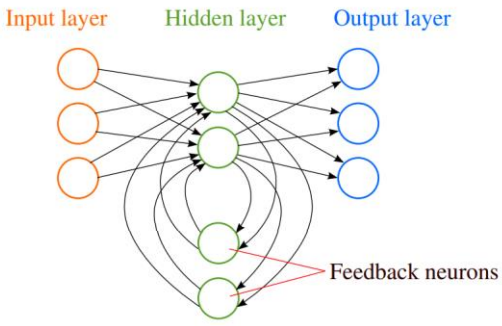
A Figura 10 exemplifica uma Rede Neural do tipo *Feed-forward*. A rede é constituída por diversos neurônios organizados em camadas. A camada das entradas (*Input layer*) é representada por círculos verdes. Há camadas intermediárias ocultas (*hidden layers*) marcadas por círculos roxos, ou seja, os neurônios. A camada das saídas (*output layer*) é ilustrada por círculos rosas. As entradas são ligadas aos neurônios através de setas, podendo haver camadas ocultas com retroalimentação, em que as saídas dos neurônios intermediários passam por outras camadas ocultas de neurônios como entradas. Assim, geram saídas mais elaboradas ao final do processamento, dado o aprendizado da rede. Cada entrada ligada ao neurônio possui um peso associado. No caso de aplicações de NLP que utilizam redes neurais, estruturas linguísticas podem ocupar o lugar das entradas e são transformadas em valores, multiplicados por seus pesos e somados. Assim, após a aplicação de uma função não-linear a seus resultados, geram uma saída. As setas mostram as relações entre as entradas, neurônios e saídas, refletindo o fluxo de informações. O sistema aprende com base nas camadas intermediárias ocultas que reutilizam as informações injetadas nelas e, caso a saída tenha sido um erro, ainda há a possibilidade de o sistema corrigir tal erro e gerar uma saída correta ou mais adequada.

Outro tipo de rede neural são as Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs). Esse modelo é do tipo *feed-forward* e vem sendo muito utilizado no processamento e análise de imagens digitais e vídeos, processamento de língua natural (*parsing*

semântico e detecção de paráfrases) e sistemas de recomendação. A rede busca identificar traços locais indicativos em grandes estruturas, combinando-os e gerando uma representação vetorial. A partir da representação vetorial do mapeamento de traços, das análises internas feitas e da geração de subamostras, é possível que o sistema reconheça a entrada ou reproduza algo que se aproxime ou se conecte mais adequadamente à informação de entrada. Na tentativa de reduzir ao mínimo o pré-processamento, são utilizados *perceptrons* multicamada, ou seja, estruturas que funcionam como um classificador binário que mapeia uma entrada (vetor) a uma saída (vetor) através de uma matriz. Esse tipo de rede é utilizado no reconhecimento de imagens e não gera uma única saída, mas uma lista dos mais prováveis resultados possíveis.

Há também as Redes Neurais Recorrentes (*Recurrent Neural Network* - RNNs). Trata-se de poderosos algoritmos úteis no processamento de dados sequenciais, sejam eles sons, palavras etc. "[...] Uma rede recorrente pode consistir em uma única camada de neurônios, com cada neurônio retornando seu sinal de saída às entradas de todos os outros neurônios." (HAYKIN, 2008, p. 43)[26]. A saída de uma camada específica é salva para poder retornar à entrada. Esse é o mecanismo que ajuda o sistema a prever o resultado da camada. Esse tipo de rede se difere das redes *feed-forward*, pois envolve um *loop* de *feedback*, ou seja, a saída do passo n-1 alimenta a rede de volta, interferindo no resultado do passo n. Se pensarmos que a exposição de dados à rede é uma palavra letra por letra, e pede-se que a rede adivinhe a próxima letra, a primeira letra contribuirá para que a rede neural recorrente preveja e sugira a próxima letra, visto que informações oferecidas pela entrada são armazenadas. Vejamos na Figura 11 a representação de uma Rede Neural Recorrente.

Figura 11 – Exemplo de uma Rede Neural Recorrente

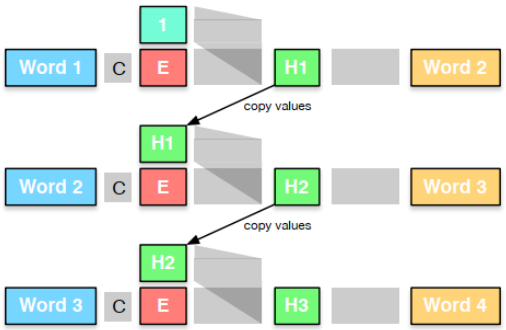


Fonte: Intelligent Predictions: an empirical study of the Cortical Learning Algorithm (GALETZKA, 2014, p. 12).

[26] "[…] a recurrent network may consist of a single layer of neurons with each neuron feeding its output signal back to the inputs of all the other neurons." (HAYKIN, 2008, p. 43).

A Figura 11 demonstra o funcionamento de uma Rede Neural Recorrente. Observamos a camada de entrada (*input layer*) representada pelos círculos laranjas, a camada de saída (*output layer*) marcada por círculos azuis e a camada intermediária oculta (*hidden layer*) ilustrada por círculos verdes. Dentro da camada intermediária oculta, há os neurônios de *feedback* (*feedback neurons*), ou seja, aqueles que guardam as informações e cálculos, retroalimentando os outros neurônios intermediários ocultos e possibilitando ao sistema aprender com os dados armazenados anteriormente. A primeira camada da rede neural recorrente funciona da mesma maneira que a rede *feed-forward*, uma vez que considera o produto da soma de pesos atribuídos às entradas. A diferença aparece nas camadas seguintes, pois cada nó na rede guardará informações e dados que possuía anteriormente. Cada neurônio intermediário funciona como uma célula de memória que computa e executa operações. Notemos na Figura 12 a representação de um modelo de língua genérico com uma arquitetura de Rede Neural Recorrente.

Figura 12 – Modelo de língua genérico por Rede Neural Recorrente



Fonte: Neural Machine Translation (KOEHN, P., 2017, p. 45).

A Figura 12 traz um modelo de língua genérico com uma arquitetura de rede neural recorrente. Essa arquitetura pode ser utilizada para diversas tarefas de NLP como a Tradução por Máquina. As palavras (*words*) em uma sentença são representadas pelos retângulos na cor azul claro, sendo que cada uma corresponde a um vetor de alta dimensão chamado *one-hot*. Por exemplo, a palavra “cachorro” pode ser expressa através de um vetor *one-hot*: cachorro = (0, 0, 0, 0, 1, 0, 0, 0, 0, ...)^{T}. Há uma matriz C (*Embedded matrix*) em um quadrado cinza representando o mesmo peso para todas as palavras. Dentro dessa matriz pode haver de 500 a 1000 nós, mas, no início, ela não impacta matematicamente em nada. Há um conceito muito importante dentro dessa arquitetura que são os *word embeddings* representados pela letra E nos quadrados na cor vermelha, ou seja, grupos de palavras com números de pontos flutuantes,

representando um vetor. Esses *word embeddings* permitem uma generalização entre palavras através de uma clusterização ou agrupamento de palavras possíveis e, assim, obter previsões a partir de contextos "invisíveis". Ainda na Figura 12, temos quadrados verde-claros com a letra H em seu interior simbolizando a palavra como o vetor de alta dimensão *one-hot*. Por fim, a palavra (*word*) seguinte gerada na sentença é ilustrada pelos retângulos na cor amarela mais à direita da representação. Devemos observar que a característica da rede neural recorrente de guardar informações acerca da palavra para serem utilizadas posteriormente no processamento é representada pelos valores de cópia (*copy values*) do vetor. Depois, são transportados para a linha debaixo, sendo incorporados às *embedded words* e facilitando a aprendizagem por parte da rede neural.

Outro conceito muito importante aplicado à tradução por redes neurais é o mecanismo de Atenção (*Attention*). Essa técnica faz com que determinadas partes da sentença-fonte sejam focalizadas durante a tradução. A partir de tal foco, o sistema consegue projetar as melhores traduções, dada a relação entre as palavras focalizadas e suas palavras-vizinhas. Luong *et al.* (2015) propõem que o mecanismo de atenção pode ser de dois tipos: o local e o global. "[...] Uma abordagem global que sempre atende a todas as palavras-fonte e uma local que analisa apenas um subconjunto de palavras-fonte por vez." (LUONG *et al.*, 2015, p. 1412)[27]. Os autores postulam que, na abordagem da atenção global, a cada etapa, o modelo deduz um vetor de peso de alinhamento de comprimento variável baseado no estado de destino atual e nos estados de origem. A partir disso, calcula-se um vetor de contexto global e computa-se a média ponderada sobre todos os estados de origem. Observamos que a sentença é considerada como um todo, bem como seu tamanho e os alinhamentos dentro dela. Já na abordagem de atenção local, os autores afirmam que o modelo prevê uma única posição de alinhamento para a palavra de destino atual. É como se uma janela focalizasse apenas na posição de origem que seria utilizada para a realização do cálculo de um vetor de contexto. A partir daí, o sistema calcularia uma média ponderada dos estados ocultos de origem na janela. Assim, os pesos seriam deduzidos a partir do estado-alvo atual e dos estados de origem realçados por essa janela. Nessa abordagem local, o sistema faz uma análise mais específica de uma determinada palavra na sentença, considerando as palavras adjacentes a ela e prevendo uma posição única de destino ao final da tradução. O mecanismo de atenção também contribui para um alinhamento entre palavras da entrada e palavras da saída.

[27] "[…] a global approach which always attends to all source words and a local one that only looks at a subset of source words at a time." (LUONG *et al.*, 2015, p. 1412).

Outro modelo bastante utilizado pela tradução por redes neurais é a Memória de Longo e Curto Prazo (*Long Short Term Memory* - LSTM). Nos sistemas de tradução por máquina em geral, a importância das palavras em relação à palavra-alvo que está sendo traduzida decai com a distância entre as palavras. Isto é, se uma palavra estiver muito afastada da palavra-alvo traduzida, é provável que sua relação com tal palavra não seja tão relevante. Koehn (2017, p.46) apresenta os seguintes exemplos para discutirmos essa relação de proximidade entre as palavras.

(1) *After much economic progress over the years, the* **country** → *has*[28]

(2) *The* **country** *which has made much economic progress over the years still* → *has*[29]

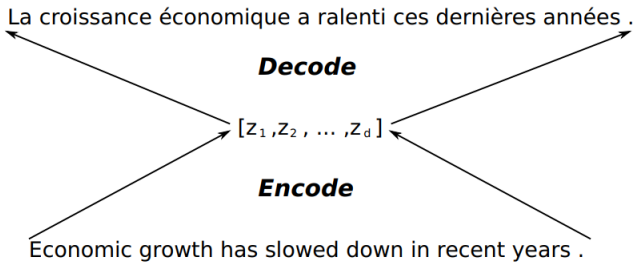
Ao abordarmos o exemplo (1), notamos que, para o sistema propor a forma verbal *has* (tem) ao invés de *have* (têm) logo após *country* (país), a palavra *country* é a mais informativa dada a distância próxima entre as duas. Entretanto, as línguas podem possuir orações subordinadas longas representadas pela sentença (2) que também funcionariam como sujeito para a formal verbal *has*, visto que o núcleo ainda continuaria sendo *country*. Assim sendo, o sistema não poderia considerar a palavra mais próxima, mas sim, detectar com qual palavra o verbo deve concordar. "Os estados ocultos (*hidden states*) nas redes neurais tendem sempre a se atualizar com a palavra mais recente, e sua memória de palavras mais antigas vai provavelmente diminuir com o tempo." (KOEHN, 2017, p. 46)[30]. Portanto, as redes neurais que utilizam LSTM são as melhores para lidar com as seguintes questões. Primeiro, a previsão das palavras de saída a partir dessas camadas ocultas que desempenham um papel importante na memória da rede. Em um segundo momento, o fato de não haver um mecanismo em outros sistemas que preste mais atenção à palavra diretamente anterior ou a um contexto mais extenso. Terceiro, o modelo pode ser treinado em longas sequências e inserir mecanismos de propagação reversa, fazendo com que o sistema percorra toda a sentença na geração da saída. Os modelos baseados em LSTM utilizam determinadas operações ou seguem certos parâmetros em seu funcionamento. Tais parâmetros podem ser chamados de portas ou portões (*gates*). O parâmetro da porta de entrada (*input gate parameter*) regula a quantidade de novas entradas que modificam o estado de memória. O parâmetro da porta de esquecimento (*forget gate parameter*)

[28] Após muito progresso econômico ao longo dos anos, o **país** → tem. (tradução nossa).
[29] O **país** que fez muitos progressos econômicos ao longo dos anos ainda → tem. (tradução nossa).
[30] "The hidden state in the recurrent neural network will always be updated with the most recent word, and its memory of older words is likely to diminish over time." (KOEHN, 2017, p. 46).

lida com a quantidade de estados de memória prévios que são mantidos ou esquecidos. Por fim, o parâmetro da porta de saída (*output gate paremeter*) controla o quão forte é o estado de memória que passa para a camada seguinte. Através de estratégias diversas, o modelo LSTM consegue lidar melhor com essas questões da distância entre as palavras a serem traduzidas e o que os estados de memória guardam em si durante o processo de tradução, além de permitir o tratamento de sentenças maiores.

Outra característica frequente nas redes neurais é um sistema de codificador-decodificador (*Encoder-Decoder*). "O codificador extrai uma representação de comprimento fixo de uma sentença de entrada de comprimento variável e o decodificador gera uma tradução correta dessa representação." (CHO *et al.* 2014, p. 103)[31]. O codificador (*Encoder*) transforma a sequência de palavras da entrada em vetores, lidos e trabalhados computacionalmente. Já o decodificador (*Decoder*) ao final do processo transforma os vetores gerados pela rede neural em sequências de palavras na saída. A Figura 13 traz uma representação superficial de uma arquitetura de rede neural com Codificador-Decodificador.

Figura 13 – Arquitetura do modelo Codificador-Decodificador



Fonte: On the Properties of Neural Machine Translation: Encoder–Decoder Approaches (CHO *et al.*, 2014, p. 105).

Na Figura 13, temos demonstrada de forma simples a arquitetura de uma rede neural que utiliza codificação e decodificação. A sentença em inglês *Economic growth has slowed down in recent years*[32] é codificada (*Encode*). Aplica-se uma distribuição condicional de probabilidades **p (f | e)** de uma sentença-alvo (tradução) **f** dada uma sentença de origem **e**. Ou seja, a partir de uma sentença **e**, qual a probabilidade de o sistema gerar uma sentença traduzida **f**? Fundamentado nessa distribuição de probabilidades, o sistema decodifica (*Decode*) a

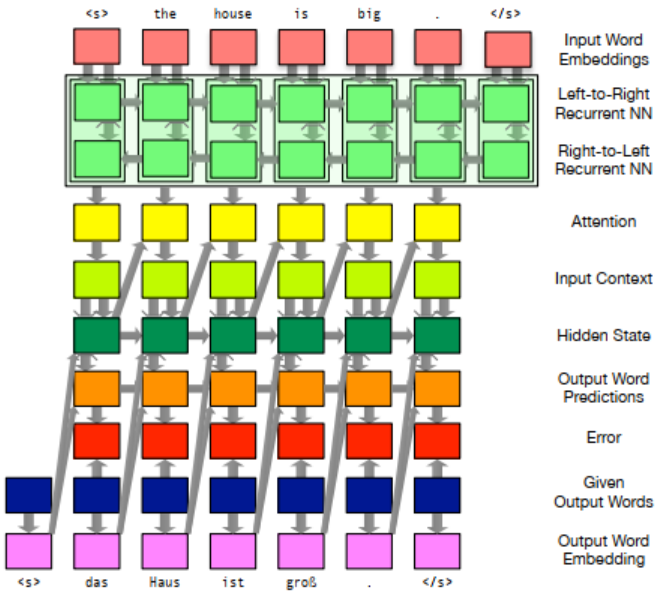
[31] "The encoder extracts a fixed-length representation from a variable-length input sentence, and the decoder generates a correct translation from this representation." (CHO *et al*. 2014, p. 103).
[32] "O crescimento econômico desacelerou nos últimos anos." (Tradução nossa).

sentença *La croissance économique a ralenti ces dernières années*[33] em francês como sendo a melhor e mais correta representação para sentença de entrada. Esse sistema considera que muitas entradas podem possuir como correspondência na língua traduzida muitas saídas. Por consequência, esse modelo torna-se melhor para lidar com questões linguísticas como a tradução por máquina, que pode apresentar sequências longas de palavras na língua-fonte a serem traduzidas em sequências longas de palavras na língua-alvo. Portanto, trata-se de um modelo de tradução muitas-para-muitas e não aqueles modelos que consideravam apenas a tradução de palavra-por-palavra. A seguir demonstramos de que modo esses conceitos se aplicam em um modelo neural de tradução por máquina.

2.4.2 Modelos Neurais de Tradução por Máquina

A Figura 14 traz uma representação mais específica e completa de todas as etapas de um modelo de tradução por máquina baseado em Redes Neurais Recorrentes.

Figura 14 – Representação de todas as etapas de um modelo de tradução por máquina baseado em Redes Neurais Recorrentes



Fonte: Neural Machine Translation (KOEHN, P., 2017, p.59).

A Figura 14 ilustra de forma mais completa uma Rede Neural Recorrente aplicada à TM. Acima da primeira camada temos exemplificada a sentença em inglês *the house is big.* (A

[33] "O crescimento econômico desacelerou nos últimos anos." (Tradução nossa).

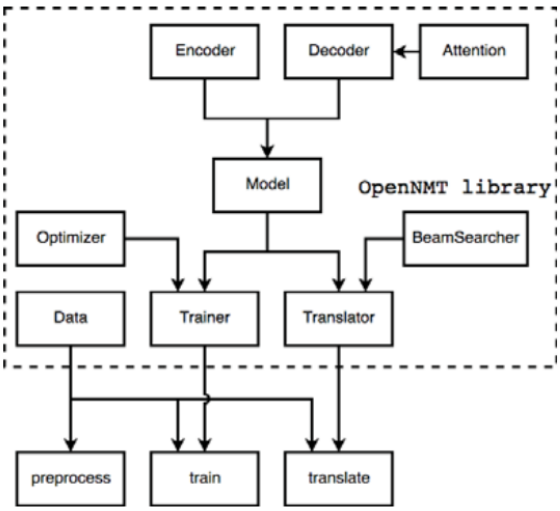
cada é grande) a ser traduzida do inglês para o alemão. Os símbolos <s> no início e </s> no final marcam computacionalmente o início e final da sentença, respectivamente, para critérios de reconhecimento de seus limites. Na primeira camada temos os *word embeddings* de entrada (*Input Word Embeddings*), ou seja, um agrupamento de palavras possíveis, em que cada palavra se encaixa em um dos retângulos na cor salmão na primeira camada. A segunda e a terceira camadas com retângulos na cor verde claro representam os nós intermediários da Rede Neural Recorrente da esquerda para a direita e vice-versa. Essa análise bidirecional marca a fase de codificação da sentença e de seus componentes (palavras e pontuação), permitindo uma análise de contexto tanto das palavras posteriores ou anteriores a que se quer traduzir. Em seguida, surge o mecanismo de Atenção (*Attention*) simbolizado pelos retângulos amarelos. Ele contribui para a análise dos elementos vistos de forma global e de forma localizada. Considera-se então o Contexto de Entrada (*Input Context*) marcado pelos retângulos na cor verde-limão. Passa-se então aos Estados Ocultos (*Hidden State*) na cor verde escuro, que guardaram as informações prévias no sistema. O sistema realiza as previsões de palavras de Saída (*Output Word Predictions*) nos retângulos alaranjados com base em toda a análise feita pela rede neural. A função de custo ou de erro é calculada para cada saída individualmente e retomada na sentença como um todo. Ela pode ser simbolizada pela Camada de Erro (*Error*) nos retângulos vermelhos. Por fim, temos a camada de Palavras de Saída Geradas (*Given Output Words*) pela rede neural como tradução para a sentença e marcadas pelos retângulos na cor azul escuro. Representando a fase de decodificação (*Decoder*), o sistema pode recorrer a estados anteriores marcados pelas setas que voltam a algumas camadas e, usando todo o contexto e as estratégias aplicadas pela rede neural recorrente, ele propõe a Camada de saída (*Output Word Embedding*) nos retângulos da última camada. Essa camada pode ser interpretada como a distribuição de probabilidades de que tais palavras irão ocorrer dentro de tais contextos específicos, estando cada sugestão de correspondente de tradução representado individualmente por um retângulo rosa. Abaixo do gráfico, temos a sentença em alemão gerada como equivalente de tradução em *das Haus ist groß*. (A cada é grande). Cada palavra da sentença e os sinais de pontuação são representados por retângulos na cor rosa, individualmente.

A título de exemplo, passemos ao funcionamento do Google Tradutor, ferramenta atual que utiliza redes neurais e apresenta os melhores resultados como equivalentes de tradução. Os modelos de Rede Neural Recorrente utilizados pelo Google tradutor são do tipo LSTM. O modelo possui três etapas: a rede de codificação, a rede de decodificação e a rede de atenção. Há oito camadas de codificação e oito camadas de decodificação, havendo conexões residuais entre as camadas, o que incentiva o fluxo de informações. A camada de Atenção do

decodificador inferior da rede se liga à camada superior codificadora da rede. Para as palavras raras, Wu *et al.* (2016) afirmam que o sistema faz uso de "[...] unidades de subpalavra (também conhecidas como "pedaços de palavras") para entradas e saídas no sistema" (WU *et al.*, 2016, p. 2)[34]. Com a utilização dessa estratégia, há um equilíbrio entre caracteres únicos e palavras completas na fase de decodificação, evitando-se a necessidade de um tratamento especial para as palavras raras ou desconhecidas. Ocorre também a técnica de Busca em Feixe (*Beam Search*), havendo uma normalização de tamanho da entrada, e uma penalidade de cobertura (*coverage penalty*), ou seja, o sistema é forçado a tentar cobrir mais material textual quando aplicada essa penalidade, o que faz com o que modelo traduza toda a entrada fornecida. Caso o modelo não consiga uma tradução adequada, ele copia a entrada e a replica na saída. Concluindo, essa abordagem se diferencia das restantes em TM por redes neurais, pois busca resolver os problemas mencionados anteriormente como as palavras raras, treinamento lento, inferência ineficaz e os problemas de não traduzir todas as palavras da entrada através das técnicas que acabamos de descrever.

Apresentaremos agora o OpenNMT, um framework para treinamento de uma rede neural. Para entendermos o funcionamento do sistema de tradução aplicado no OpenNMT, observemos a Figura 15.

Figura 15 – Visão geral esquemática do código OpenNMT-Python



Fonte: OpenNMT: Neural Machine Translation Toolkit (KLEIN, G. *et al.*, 2018, p.180).

[34] "[…] sub-word units (also known as "wordpieces") for inputs and outputs in our system." (WU *et al.*, 2016, p. 2).

Na Figura 15, temos ilustrado um esquema do funcionamento da rede neural do OpenNMT voltado à TM. Na parte inferior do gráfico, notamos as três etapas básicas de funcionamento da rede neural: pré-processamento (*preprocess*), treinamento (*train*) e tradução (*translate*). Dentro do retângulo tracejado encontram-se as etapas internas da rede neural que dão suporte a cada uma dessas etapas gerais da rede neural voltada à TM. O pré-processamento é onde entram os dados (*data*). Eles podem ser tratados previamente antes de serem inseridos no sistema. Na etapa de treinamento, a rede neural é treinada, aprende e só a partir daí ela está pronta para traduzir as palavras ou sentenças. O modelo (*model*) funciona a partir de um corpus de treinamento com as etapas de codificação (*encoder*), decodificação (*decoder*) e o mecanismo de atenção (*attention*), todos previamente explicados neste trabalho. Um otimizador (*optmizer*) é aplicado ainda na etapa de treinamento, visando a melhorar possíveis deficiências no funcionamento. O treinamento da rede pode demorar dias a depender de quantas unidades de processamento gráfico (GPUs) estão sendo utilizadas. A estratégia de Busca em Feixe (*Beam Search*) é inserida na etapa da tradução e faz uso de dados de uma biblioteca do OpenNMT. Lembrando que a busca em feixe contribui para o sistema com estratégias de alinhamento. A própria ferramenta oferece corpora para que a rede seja treinada e em seu código permite que sejam extraídas informações de outras bases de dados, caso seja necessário.

Os sistemas de tradução por redes neurais têm sido atualmente os melhores recursos aplicados na área de TM. Entretanto, ainda apresentam alguns problemas. Wu *et al.* (2016) apontam como problemas de sistemas de TM por redes neurais "[...] seu treinamento mais lento e velocidade de inferência, ineficácia ao lidar com palavras raras e, às vezes, falha na tradução de todas as palavras na sentença de origem." (WU *et al*., 2016, p. 2)[35]. Passemos agora à seção 2.5 e os modelos híbridos de TM, que buscam solucionar alguns desses problemas.

2.5 MODELOS HÍBRIDOS DE TM

A tradução por máquina híbrida (*Hybrid Machine Translation*) pode ser compreendida como sistemas de tradução por máquina que integram arquiteturas variadas de TM em um mesmo sistema. Essa hibridização é motivada pela possibilidade de correção de falhas em uma das abordagens utilizadas com características da(s) outra(s), buscando traduções mais adequadas e um maior nível de precisão. Kamran (2013) apresenta uma separação dos modelos

[35] "[…] its slower training and inference speed, ineffectiveness in dealing with rare words, and sometimes failure to translate all words in the source sentence." (WU *et al.*, 2016, p. 2).

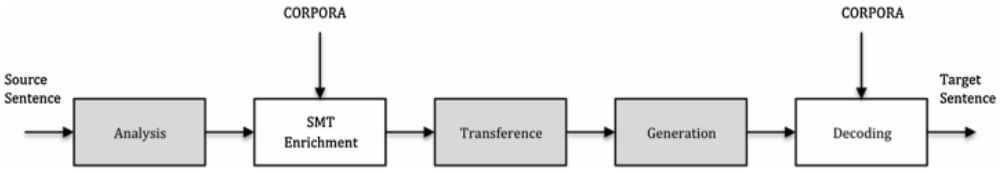
de hibridização em duas categorias: a Hibridização em um Único Sistema (*Single-Engine Hybridization* – SEH) e a Hibridização em Multisistemas (*Multi-Engine Hybridization* – MEH).

A Hibridização em um Único Sistema possui uma arquitetura principal fundamentada em apenas uma abordagem, ou a baseada em regras (*Rule-based Machine Translation* – RBMT), ou a estatística (*Statistical Machine Translation* – SMT), sendo esta uma subcategoria de Sistemas baseados em Corpus (*Corpus-based*). As modificações podem ser inseridas no sistema em diversos níveis a depender dos problemas que pretendem resolver ou de questões que pretendem aprimorar na TM. Como exemplos dessa integração, mencionemos um sistema que utiliza RBMT, mas modificado por técnicas estatísticas diversas, ou um outro sistema que utiliza como base SMT, mas incorpora informações linguísticas. Costa-Jussà e Fonollosa (2015) esclarecem que

> [...] as abordagens de TM baseadas em regras (ou seja, RBMT) usam informações linguísticas, como dicionários monolíngues e bilíngues, combinadas com o conhecimento linguístico humano. As regras são desenvolvidas manualmente para transferir texto no idioma de origem para um texto no idioma de destino. As abordagens mais populares do RBMT aplicam três fases diferentes: análise, transferência e geração. (COSTA-JUSSÁ; FONOLLOSA, 2015, p. 4)[36].

A partir da exposição feita por Costa-Jussà e Fonollosa acerca do funcionamento de um sistema de TM baseado em regras genérico e suas etapas, vejamos na Figura 16 a representação da hibridização em um sistema de TM baseado em regras com a incorporação de técnicas estatísticas no modelo.

Figura 16 – Esquema de hibridização de TM baseada em regras



Fonte: Latest trends in hybrid machine translation and its applications (COSTA-JUSSÁ, M.; FONOLLOSA, J., 2015, p. 5).

A Figura 16 exibe um esquema de TM híbrida baseado em regras (RBMT) e enriquecido com técnicas estatísticas. A sentença de entrada ou sentença-fonte (*source sentence*) é

[36] "MT approaches based on rules (i.e. RBMT) use linguistic information such as monolingual and bilingual dictionaries combined with human linguistic knowledge. Rules are developed manually to transfer text in a source language text into a target language text. Most popular RBMT approaches apply three different phases: analysis, transfer and generation." (COSTA-JUSSÁ; FONOLLOSA, 2015, p. 4).

submetida a uma fase de análise. Nessa fase de análise (*analysis*), são utilizados os dicionários e gramáticas da língua-fonte para que seja realizado um *parsing* ou rotulação sintática, vinculando as palavras a categorias sintáticas da língua-fonte. Na fase de transferência (*transference*), o sistema faz substituições lexicais a partir do pareamento com a estrutura da língua de destino. Por fim, a fase de geração (*generation*) realiza um certo alinhamento e oferece a tradução da sentença-fonte na língua de destino. No entanto, o sistema híbrido de tradução incorpora características estatísticas em duas etapas. Logo após a fase de análise e do *parsing*, a TM estatística aparece como uma forma de enriquecer (SMT *Enrichment*) a rotulação atribuída às palavras da sentença-fonte a partir de informações compiladas de grandes corpora na língua-fonte. Essa característica contribui para que o sistema consiga mapear construções e padrões da língua-fonte, trazendo um refinamento linguístico para a rotulação sintática atribuída às palavras da sentença-fonte. Após a fase de geração da RBMT, entraria mais uma etapa estatística de decodificação (*decoding*), em que o sistema utilizaria também grandes quantidades de corpora na língua de destino para que a decodificação das palavras e a ordem de alinhamento delas na sentença-alvo (*target sentence*) traduzida apresentem maior semelhança com a estrutura e usos da língua de destino.

Há também os sistemas híbridos de tradução que se fundamentam na TM Estatística e realizam modificações em certas etapas através de técnicas da RBMT, não requerendo inicialmente muito conhecimento linguístico. Karlbom (2016) explica que a ideia nos modelos estatísticos de TM é que o sistema busca,

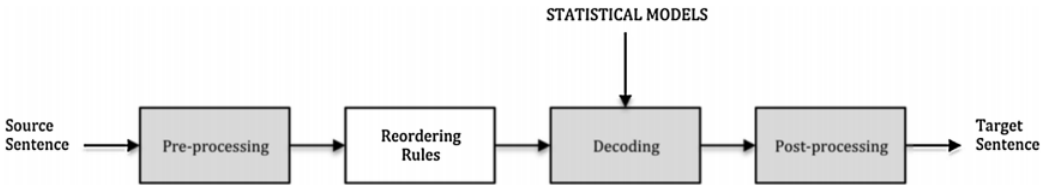
> [...] dada uma sentença na língua-fonte, encontrar a tradução para a língua-alvo que tenha maior probabilidade. Há três etapas em que os sistemas SMT estão preocupados: modelagem, treinamento e decodificação. A modelagem orienta-se a definir um método para calcular a probabilidade de uma sentença na língua-alvo a partir de uma sentença na língua-fonte. A etapa de treinamento debruça-se sobre a utilização de um corpus para estimar os parâmetros do modelo que foi definido. Por fim, o problema de decodificação se concentra na busca para encontrar a sentença que tenha a maior probabilidade entre todas as candidatas de tradução. (KARLBOM, 2016, p. 11)[37].

Há duas possibilidades de incorporação de técnicas de RBMT nos sistemas híbridos de SMT. A primeira é através do uso de regras em etapas de pré-processamento e pós-edição. A

[37] "The main is idea is, given a sentence in the SL, to find the translation to the TL that has highest probability. There are three steps which SMT systems are concerned with: modeling, training and decoding. The modeling is about how to define a method for calculating the probability of a TL sentence having a SL language sentence. The training step is to use a corpus to estimate the parameters of the model which was defined. Lastly the decoding problem focuses on the search to find the sentence which has the highest probability among all the candidate translations." (KARLBOM, 2016, p. 11).

segunda ocorre pela integração de dicionários ou regras dentro do modelo. A partir do resumo das etapas básicas de SMT proposto acima por Karlbom (2016), sejam elas modelagem, treinamento e decodificação, vejamos na Figura 17 o esquema de TM híbrido baseado em estatística que incorpora técnicas da RBMT.

Figura 17 – Esquema de hibridização de TM baseada em estatística



Fonte: Latest trends in hybrid machine translation and its applications (COSTA-JUSSÁ, M.; FONOLLOSA, J., 2015, p. 6).

Na Figura 17, notamos que a sentença-fonte (*source sentence*) passa por uma fase de pré-processamento (*pre-processing*). Os sistemas de SMT normalmente possuem essa etapa. Nela, os dados são transformados a fim de serem utilizados pelo sistema, podendo passar por transformações na forma e no tamanho. Um exemplo de transformação são as formas contraídas do inglês como *aren't* (não estão/não são), que é transformada em *are not* com o intuito de contribuir para que o sistema reconheça as estruturas e não perca informação. As ferramentas de pré-processamento podem funcionar a partir de regras escritas por humanos, extraídas automaticamente de textos, regras sintáticas das línguas, entre outras. O sistema então é treinado utilizando modelos estatísticos (*statistical models*). A partir desse treinamento baseado em corpora e nos parâmetros estipulados previamente na modelagem, o sistema decodifica (*decoder*) qual a sentença-alvo tem a maior probabilidade de ser o melhor equivalente de tradução da sentença-fonte.

Com o propósito de resolver problemas encontrados na tradução gerada, sejam de alinhamento, concordância nominal ou verbal, entre outros, há uma etapa de pós-processamento (*post-processing*) antes de o sistema oferecer a sentença-alvo (*target sentence*) final equivalente de tradução. Nessa etapa de pós-processamento podem ser inseridos dicionários e gramáticas, também contribuindo para que o sistema passe a possuir as variações morfológicas da língua de destino, gerando traduções mais adequadas com informações linguísticas que não foram ressaltadas no corpus de treinamento. A técnica de RBMT que pode ser inserida nesse sistema híbrido de SMT é composta por as regras de reordenamento (*reordering rules*) aplicadas logo após o pré-processamento. Essas regras trariam a um sistema estatístico de TM mais

informações linguísticas aplicadas à sentença pré-processada, o que poderia contribuir para um melhor produto ao final do processo.

Esta tese se propõe à utilização de um sistema de tradução por máquina híbrido baseado em redes neurais através da utilização do Sistema NMT do Google Tradutor V2 que utiliza *Transformers* Universais, ou seja, uma nova arquitetura de RNN com um mecanismo de autoatenção. Há uma ferramenta de desambiguação dos frames evocados pelas ULs do texto de referência e do texto traduzido (DAISY, sigla em inglês para *Disambiguation Algorithm for Inferring the Semantics of Y*), que utiliza as relações entre frames e as relações qualia em seu processo de desambiguação. A DAISY pode ser executada tanto na fase de pré-processamento, quanto na fase de pós-edição. Na etapa de pós-edição de um dos sistemas neurais híbridos de TM desenvolvidos nesta tese, o material traduzido também é submetido a processos de injeção terminológica e substituição lexical, o que traz um aspecto inovador para este trabalho, dado o fato de as etapas de pré-processamento e pós-edição proporcionarem um refinamento semântico nos dados. Mais detalhes sobre este sistema serão apresentados na seção 6.1.

Passando à segunda forma de incorporação de técnicas de RBMT nos sistemas híbridos de SMT, mencionemos a incorporação de dicionários e/ou regras no interior do modelo principal. Essa integração pode ocorrer com a inserção de morfologia e sintaxe nas informações extraídas de corpora. Os sistemas também podem ter módulos baseados em sentenças prontas que contribuem para um melhor alinhamento. Eles podem ainda realizar consultas em tabelas de pares de frases e suas traduções e, a partir disso, gerar melhores traduções, dado o fato de esses pares terem sido curados por tradutores humanos.

O outro tipo de categoria de sistema híbrido de TM é a Hibridização em Multisistemas (*Multi-Engine Hybridization*). Essa arquitetura, também chamada de Acoplamento (*Coupling*), combina dois ou mais sistemas de TM existentes visando a gerar saídas mais aprimoradas. A Hibridização em Multisistemas divide-se em duas abordagens: uma simples e uma mais sofisticada. Na abordagem simples, seleciona-se a melhor saída gerada por vários sistemas, deixando as hipóteses individuais como estão. Trabalha-se com *n-grams* e frases gerados pelos diferentes sistemas e não com a sentença toda. Ao final do processo, é possível a junção das melhores frases traduzidas por sistemas diversos, compondo uma sentença traduzida de forma mais adequada do que se as sentenças fossem totalmente traduzidas por um único sistema. A outra abordagem é mais complexa e elaborada. Nessa arquitetura, há uma recombinação das melhores saídas geradas por diferentes sistemas. Novamente, essa abordagem não lida com as sentenças inteiras, mas com segmentos menores como frases e palavras. A recombinação de partes pode acarretar melhores saídas do que aquelas geradas por sistemas únicos.

Como exemplo de aplicação de sistemas híbridos de TM, citemos o PROMT (*PROject Machine Translation*). Esse sistema de TM começou a ser desenvolvido inicialmente em 1991. Essa abordagem se fundamenta tradicionalmente na RBMT. O sistema emprega amplas bases de dados linguísticos que contêm traços morfológicos, lexicais e sintáticos voltados para algumas línguas tais como o inglês, alemão, francês, português, espanhol, italiano, russo, entre outras. Atualmente, o sistema considera 51 pares de línguas partindo de 13 línguas-fonte, além do uso de dicionários bilíngues com até 250 mil entradas para cada par de línguas. Molkanov e Bykov (2016) esclarecem que o sistema de tradução do PROMT possui um pré-processamento que filtra dados em paralelo, remove sentenças muito longas e coloca todas as palavras em caixa-baixa (letras minúsculas). Há também um processamento de Entidades Nomeadas (NEs) utilizando XML (*Extensible Markup Language*) que marca e guarda as características especiais que as NEs apresentam, realizando o processamento múltiplo de nomes de pessoas e de empresas, números de telefone, e-mails e datas. Molkanov e Bykov (2016) explicam que o PROMT possui um sistema híbrido de TM composto por três componentes que trabalham na tradução do texto, sejam eles: um módulo RBMT, um pós-processador RBMT e o módulo de pós-edição estatística (SPE).

> Primeiro, o módulo RBMT traduz o texto-fonte e gera uma estrutura complexa que contém a tradução e suas características linguísticas (informações morfológicas e sintáticas, entidades nomeadas extraídas etc.). Em um segundo momento, o pós-processador RBMT gera o XML com base na saída do módulo RBMT. Finalmente, o XML é alimentado no módulo SPE, que gera a conversão de saída. O módulo SPE é basicamente um sistema SMT construído em um corpus paralelo de traduções RBMT e suas referências humanas. (MOLCHANOV; BYKOV, 2016, p. 339)[38].

A inserção de uma etapa de pós-edição estatística possibilita ao sistema lidar com erros sistemáticos que a abordagem baseada em regras possa vir a possuir, além de contribuir para uma tradução mais rápida e precisa de domínios específicos, dado o fato de poderem lidar com corpora e grandes quantidades de dados estatísticos de domínio específico. Molchanov (2018) salienta que atualmente o PROMT está em funcionamento com três tipos de sistemas: um puramente neural, um híbrido RBMT com um módulo neural de pós-edição, e um puramente RBMT. O sistema puramente neural lida com o *toolkit* Marian, possuindo um módulo de

[38] "First, the RBMT module translates the source text and outputs a complex structure containing the translation and its linguistic features (morphological and syntactic information, extracted named entities etc.). Second, the RBMT postprocessor generates XML based on the output of the RBMT module. Finally, the XML is fed to the SPE module, which generates the output translation. The SPE module is basically a SMT system built on a parallel corpus of RBMT translations and their human references." (MOLCHANOV; BYKOV, 2016, p. 339).

processamento de entidades nomeadas RBMT e um mecanismo de retorno de informações ao sistema também baseados em regras. O sistema híbrido baseado em Tradução Neural é fundamentado em um sistema baseado em regras, mas que possui um módulo de pós-edição neural que utiliza o Sistema NMT do Google Tradutor V2 que utiliza *Transformers* Universais. O último sistema que seguem utilizando é um puramente baseado em regras.

Percebemos que os sistemas híbridos oferecem uma alternativa significativa para a mesclagem das melhores características de sistemas diversos, além da possibilidade de incorporação de novas ferramentas ou traços no pré-processamento ou pós-edição do sistema, o que contribuiria para a resolução de determinados problemas que uma ou outra abordagem de TM não dão conta de forma satisfatória. Passemos agora para a descrição de métodos de pré-processamento e pós-edição em TM na seção 2.6.

2.6 MÉTODOS DE PRÉ-PROCESSAMENTO E PÓS-EDIÇÃO

O pré-processamento e a pós-edição são etapas importantes que podem existir em sistemas de TM. Elas podem ocorrer de forma humana manual, com a utilização de *softwares* de apoio à tradução (CATs) ou de forma automática através de ferramentas computacionais. Simard *et al.* (2007) evidenciam em seu experimento que o processo de pós-edição pode ocorrer de forma automática e ser bem-sucedido, além de ilustrarem que técnicas de pós-edição podem oferecer melhor qualidade e detalhamento ao texto traduzido.

Por pré-processamento entende-se qualquer modificação realizada na sentença-fonte que será submetida a um sistema de TM a fim de colaborar para que o sistema possa desempenhar o processo tradutório de forma mais produtiva. Boillon *et al.* (2014) apontam que o pré-processamento pode ser empregado nas seguintes formas: verificação de gramática e ortografia, reordenamento, normalização lexical, Língua Natural Controlada. Como exemplo de verificação ortográfica, destacam-se as palavras escritas de forma incorreta, os erros de digitação como letras trocadas e a falta ou excesso de letras em determinadas palavras. Passando à verificação gramatical, mencionemos problemas de concordância nominal e verbal, modo verbal, uso de verbos impessoais, regência adequada, verbos de duplo particípio, pronomes pessoais oblíquos e grau dos adjetivos e advérbios, por exemplo. Já o reordenamento inclui a reorganização das sentenças de modo a facilitar a tradução levando em conta as diferentes estruturas de sentença observadas entre as línguas. Um exemplo seria a transformação de sentenças que estejam na ordem não canônica da língua para a ordem canônica, a fim de contribuir para que o sistema não gere problemas de reordenamento posteriormente. Por

normalização lexical entende-se uma estratégia de adequação lexical para que as sentenças possuam a menor quantidade de itens fora do vocabulário (do inglês, *Out-Of-Vocabulary*, OOV) quando submetidas ao sistema de TM. Banerjee *et al.* (2012) ilustram os tipos de técnicas de normalização lexical e apresentam que as mais comuns são: (i) a adequação de data, hora, locais, endereços; (ii) a transcrição de números em algarismo para sua versão por extenso; (iii) a separação de palavras fundidas em que dois *tokens* se mesclaram, correção e adequação de erros de ortografia e erros tipográficos; (iv) o reconhecimento, como válidas, de palavras que não foram validadas anteriormente por não estarem presentes no corpus de treinamento; e, por fim, (v) a ocorrência de palavras que indicam nomes próprios de locais, produtos e serviços que não deveriam ser traduzidos por serem utilizados em sua versão original independentemente da língua em que forem empregados. Como última forma de pré-processamento, mencionemos a definição dada por Gomes (2010) para a Língua Natural Controlada (*Controlled Natural Language* – CNL).

> Uma linguagem controlada não é, pois, uma linguagem artificial, mas uma forma controlada/simplificada da linguagem natural por meio de regras gramaticais e de um vocabulário reduzido e normalizado. Distinguem-se aqui as linguagens controladas das línguas da especialidade. Estas últimas restringem-se a uma área do saber – como por exemplo, a Medicina, o Direito, a Linguística, entre outras -, tendo, por conseguinte, um vocabulário particular. (GOMES, 2010, p. 37).

Com a definição de CNL proposta por Gomes (2010), enfatizemos o Português Controlado. Nas línguas controladas há uma restrição de gramática e vocabulário, sendo mais simples, reduzidas e contribuindo para a eliminação de ambiguidades da língua. São dois tipos de línguas controladas: aquelas que contribuem para uma melhor legibilidade por leitores humanos, como, por exemplo, textos adaptados para aprendizes não-nativos de uma língua estrangeira, e as línguas controladas que colaboram para que o trabalho dos *parsers* e de analisadores semânticos sejam melhor desempenhados.

Passando à pós-edição (*Post-Editing Machine Translation* - PEMT), trata-se de um método que vem sendo utilizado em diversas áreas de NLP, tais como a correção automática de erros, o reconhecimento ótico de caracteres, as memórias de tradução, a língua natural controlada, além de ser uma etapa possível aplicada em sistemas de TM. Allen (2003) resume o conceito de pós-edição, afirmando que "[...] a tarefa do pós-editor é editar, modificar e/ou corrigir o texto pré-traduzido que foi processado por um sistema TM de um idioma de origem

para o(s) idioma(s) de destino." (ALLEN, 2003, p. 297)[39]. A pós-edição tende a ser aplicada na tentativa de se minimizarem os erros ou problemas encontrados após o sistema de TM traduzir a sentença-fonte, desempenhando adequações e melhorias na tradução realizada. A pós-edição normalmente é feita por humanos, embora atualmente, os modelos de TM permitam a inserção de ferramentas ou etapas em sistemas de TM que oferecem um tratamento linguístico-computacional (morfológico, sintático, semântico) ao segmento traduzido pelo algoritmo.

A pós-edição automática (*Automatic Post-editing* – APE) desempenha tarefas específicas no texto traduzido. Dougal (2018) aponta duas operações terminológicas comumente utilizadas na pós-edição: a validação e a substituição. A validação ocorre através da verificação dos termos utilizados, se são apropriados ou não dentro de uma terminologia específica. Já a substituição acontece quando se substitui um termo não aprovado ou se injeta um termo específico que não se manifestou no texto traduzido. Para que ambas sejam bem sucedidas, uma modelagem de domínio específico ou uma ontologia devem ser previamente estabelecidas.

O algoritmo pode ser programado para inserir, trocar ou apagar palavras ou frases conforme a necessidade ou especificidade apontada no texto. O gênero textual ou o fato de o texto pertencer ou possuir muitos termos de um domínio específico (Turismo, Esportes, Direito, Engenharia, Medicina etc) podem ser indícios que apontam para essa necessidade de uma substituição lexical na pós-edição. A Substituição Lexical (*Lexical Substitution*) pode ser incorporada através de algumas maneiras. Vejamos quatro dessas formas. Wicentowski *et al.* (2010) utilizam dicionários bilíngues como recurso de substituição lexical em seus experimentos. Já Germann (2014) propõe outro recurso de armazenamento de informações que podem ser utilizadas na modificação lexical. O autor salienta que "[...] as opções de tradução para as frases de origem são convencionalmente armazenadas em uma tabela pré-computada, chamada de tabela de frases" (GERMANN, 2014, p. 3)[40]. Também podem ser utilizados Trios Artificiais (*Artificial Triplets*). Junczys-Dowmunt e Grundkiewicz (2016) propõem que haja um modelo de treinamento preexistente que faz um pareamento triplo contendo uma sentença-fonte, sua tradução por máquina e sua versão pós-editada por humanos. O treinamento do sistema com esses trios artificiais contribui para a substituição lexical posterior, mas mostra-se como um recurso limitado, dada a dependência de humanos para a pós-edição ou de textos que

[39] "[…] the task of the post-editor is to edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)." (ALLEN, 2003, p. 297).
[40] "Translation options for source phrases are conventionally stored in a pre-computed table, which is called the phrase table." (GERMANN, 2014, p. 3).

foram pós-editados por humanos anteriormente. Por último, tratemos da Injeção Terminológica (*Terminology Injection*). Chatterjee *et al*. (2017) apresentam essa ferramenta e opção de substituição lexical afirmando que

> [...] a injeção de conhecimento externo no decodificador é geralmente tratada com a chamada marcação XML, uma técnica usada para guiar o decodificador, fornecendo a tradução desejada para algumas das frases de origem. A opção de tradução fornecida pode ser injetada na saída usando estratégias diferentes, todas bastante diretas. (CHATTERJEE *et al.*, 2017, p. 158)[41].

A injeção terminológica pode ser compreendida como a inserção ou modificação lexical ocorrida no texto traduzido durante a pós-edição. Esse procedimento pode ocorrer com a utilização de bases de dados externas ao sistema de TM que são incorporadas a ele. No âmbito desta pesquisa, utilizamos a injeção terminológica estruturada no pré-processamento em um dos sistemas de TM e na pós-edição no outro. Os dados lexicais injetados no sistema de TM são constituídos a partir da modelagem de frames, relações frame-a-frame e relações qualia modeladas na FrameNet Brasil, além das unidades lexicais de domínio específico do Turismo e dos Esportes também na FrameNet Brasil, modeladas e ligadas em português, inglês e espanhol.

A seguir, vejamos na seção 2.7 como funciona a Avaliação de TM e as métricas desenvolvidas como parâmetros comparativos e avaliativos de TM.

2.7 AVALIAÇÃO DE TM E MÉTRICAS

Com o passar dos anos, a tradução por máquina foi se desenvolvendo através de arquiteturas, técnicas e algoritmos diversos como vimos nas seções anteriores. Com esse avanço, a qualidade da TM passa a ser questionada e surgem formas de avaliação dessa qualidade, ou seja, métricas de avaliação de TM[42]. Trata-se de medidas ou parâmetros tomados como resultados ideais de tradução e com os quais são comparadas as saídas de um sistema de TM.

Para se avaliar a qualidade de forma objetiva, Han (2018) aponta alguns critérios de avaliação: inteligibilidade (*intelligibility*), fidelidade ou acurácia (*fidelity/accuracy*), fluência

[41] "[…] the injection of external knowledge in the decoder is usually handled with the so-called XML markup, a technique used to guide the decoder by supplying the desired translation for some of the source phrases. The supplied translation choice can be injected in the output by using different strategies, all rather straightforward." (CHATTERJEE *et al.*, 2017, p. 158).

[42] Nesta tese, métrica e medida de avaliação de TM são sinônimos.

(*fluency*), adequação (*adequacy*) e compreensão ou informatividade (*comprehension/informativeness*). Com inteligibilidade, o autor se refere à característica de que "[...] a tradução deve ser lida como prosa normal, bem editada e ser facilmente compreensível da mesma maneira que tal sentença seria compreensível se originalmente composta na língua da tradução" (HAN, 2018, p. 2)[43]. Já fidelidade ou acurácia é a capacidade do texto traduzido em manter, com o mínimo de distorções, o significado proposto pelo texto original ou pelas especificações linguísticas submetidas ao sistema de TM. A fluência indica a característica do texto traduzido de ser bem formado e fluente. A adequação diz respeito ao fato de o texto traduzido se adequar e manter as características de um contexto específico. Por fim, a compreensão ou informatividade é "[...] a capacidade de um sistema produzir uma tradução que transmita informações suficientes, para que as pessoas possam obter/compreender as informações necessárias" (HAN, 2018, p. 3)[44].

Inicialmente, as métricas possuíam o apoio de tradutores humanos, seja para a curadoria do que havia sido proposto pela tradução ou para a tradução de um texto usado como referência. Han (2018) aponta alguns problemas com a avaliação humana manual nas métricas. Para o autor, "[...] a avaliação manual sofre algumas desvantagens, como o fato de ser demorada, cara, não ajustável e não reproduzível. Devido às fraquezas nos julgamentos humanos, as métricas de avaliação automática têm sido amplamente usadas para tradução por máquina." (HAN, 2018, p.4)[45]. As métricas automáticas propiciam um padrão de avaliação seguindo critérios mais objetivos, enquanto que as avaliações com o apoio de tradutores humanos estão expostas à subjetividade e à não padronização. Exploremos então algumas propriedades das métricas e abordemos as métricas mais conhecidas de avaliação de TM.

Como propriedades principais e comuns a muitas métricas, citemos a precisão (*precision),* a revocação (*recall)* e a média harmônica *F-measure*. A precisão (*precision*) é uma propriedade que corresponde à fração de instâncias recuperadas que são relevantes, ou seja, o número de elementos recuperados que estão corretos dentro do número de elementos identificados. Sendo assim, a precisão diz respeito à quantidade de resultados gerados que foram úteis. Já a revocação (*recall)* é uma outra propriedade que equivale à fração de instâncias

[43] "[…] the translation should read like normal, well-edited prose and be readily understandable in the same way that such a sentence would be understandable if originally composed in the translation language." (HAN, 2018, p. 2).
[44] "[…] a system's ability to produce a translation that conveys sufficient information, such that people can gain necessary information from it." (HAN, 2018, p. 3).
[45] "[…] manual evaluation suffers some disadvantages such as time-consuming, expensive, not tunable, and not reproducible. Due to the weaknesses in human judgments, automatic evaluation metrics have been widely used for machine translation." (HAN, 2018, p. 4).

relevantes que são recuperadas do total que se esperava recuperar. Portanto, trata-se de quão completos são os resultados gerados. Por último, temos a *F-measure* ou *F-score*, que é a medida de acurácia do teste. Essa medida considera tanto *precision* quanto *recall*, sendo uma média harmônica de ambas. Como melhor resultado da *F-measure*, teríamos o valor 1 (um) (*precision* e *recall* perfeitos) e como pior resultado o valor seria 0 (zero). Vejamos os exemplos abaixo para compreendermos melhor essas propriedades.[46]

(3) Após soltar o disco, o lançador continua a girar.
(4) After releasing the discus, the thrower continues to turn.
(5) After releasing the record, launcher continues rotate.

Tomando uma situação em que desejamos traduzir a sentença (3) do português para o inglês, a sentença original possui 9 palavras e a sentença (4), tradução de referência, também possui 9 palavras. Já a nossa sentença (5), traduzida por máquina, possui a tradução de apenas 7 palavras, sendo que "*record"* e "*launcher*" não são as traduções mais adequadas para "disco" e "lançador" no contexto específico dos esportes. Podemos afirmar que o sistema gerou 5 palavras como traduções corretas e válidas (*after, releasing, the, continues, rotate*) em inglês das 9 palavras do total que se esperava gerar, como observado em (4). Portanto, a precisão do sistema de TM com base no teste dessa única sentença é de 5/7 ou 0,71, dado que o sistema retornou 5 resultados ou palavras traduzidas válidas e corretas de 7 traduções geradas. A revocação desse sistema será 5/9 ou 0,55, visto que o sistema retornou 5 traduções válidas e corretas de um total de 9 palavras traduzidas que se esperava retornar. Já a *F-measure* é uma medida harmônica da precisão e da revocação e pode ser computada a partir da Equação 2.

Equação 2 – Fórmula de cálculo da *F-measure*

$$F = 2 \times \frac{precisão \times revocação}{precisão + revocação}$$

Fonte: Natural Language Annotation for Machine Learning (PUSTEJOVSKY; STUBBS, 2012, p.175).

[46] O exemplo (3) foi retirado da página 64 do Livro dos Esportes (RODRIGUES; NUNO; SALERNO, 2012). O exemplo (4) é a tradução de referência em inglês, retirado da mesma página do livro em inglês The Sports Book (BRIDLE *et. al.*, 2011). O exemplo (5) trata-se de um exemplo de uma possível tradução por máquina realizada para a sentença (3).

Portanto, inserindo os dados compilados do nosso sistema-exemplo e das sentenças (3), (4) e (5), temos representada a fórmula da *F-measure* conforme vemos na Equação 3.

Equação 3 – Fórmula da *F-measure* com os dados do sistema-exemplo

$$F = 2 \times \frac{0{,}71 \times 0{,}55}{0{,}71 + 0{,}55}$$

Fonte: Elaborado pelo autor (2020).

A partir dos cálculos ilustrados na Equação 3, temos que a *F-measure* desse nosso sistema de TM é 0,61, ou seja, a média harmônica entre a precisão e a revocação do sistema.

A primeira e mais difundida das métricas de avaliação de TM é a BLEU, criada pelo grupo IBM. É apontada como rápida, barata e independente de línguas específicas, tendo como principal tarefa

> [...] comparar *n-grams* da tradução candidata com *n-grams* da tradução de referência e contar o número de correspondências. Essas correspondências são independentes da posição. Quanto mais correspondências, melhor a tradução da candidata. Para simplificar, primeiro nos concentramos em calcular correspondências *unigram*. (PAPIENI *et al.*, 2002, p. 312)[47].

A métrica BLEU é aplicada tomando inicialmente *unigrams* como unidades comparativas de tradução, ou melhor, palavras únicas são avaliadas na tradução. Há um ranqueamento das traduções candidatas em que, as que possuírem maior correspondência com o material da tradução humana de referência recebem um valor maior, sendo melhor classificadas enquanto equivalentes de tradução. A tradução de palavras únicas tende a atender melhor o critério de adequação. O processo é repetido considerando-se *bigrams*, isto é, duas palavras que ocorrem juntas, e assim por diante. As correspondências que se verificam entre *n-grams* mais longos buscam atender o critério de fluência.

Com melhorias nos critérios de avaliação, a BLEU passa a considerar um fator de penalidade multiplicativo por brevidade. Desse modo, "uma candidata de tradução com pontuação alta deve corresponder às traduções de referência em tamanho, escolha de palavras e a ordem das palavras." (PAPIENI *et al*., 2002, p. 315). Algumas limitações evidenciadas pela

[47] "[…] to compare n-grams of the candidate with the n-grams of the reference translation and count the number of matches. These matches are position independent. The more the matches, the better the candidate translation is. For simplicity, we first focus on computing unigram matches." (PAPIENI *et al*., 2002, p. 312).

BLEU são o fato de ela medir a correspondência de forma puramente lexical, não avaliar apropriadamente a adequação e não capturar as peculiaridades de significado entre a tradução e o texto de referência em textos maiores. As características sintáticas da tradução são levemente apreciadas se considerarmos o fator multiplicativo de penalidade por brevidade.

Outras métricas utilizadas na avaliação de TM são a TER e sua versão que envolve humanos no processo de edição, HTER. Seu nome diz respeito a taxas de edição de tradução por humanos (*Human Translation Edit Rate*). Snover *et al.* (2006) explicam que

> TER (Taxa de Edição de Erros de Tradução) é definida como o número mínimo de edições necessárias para alterar uma hipótese, para que ela corresponda exatamente a uma das referências, normalizada pelo comprimento médio das referências. Como estamos preocupados com o número mínimo de edições necessárias para modificar a hipótese, apenas medimos o número de edições para a referência mais próxima. (SNOVER *et al.*, 2006, p. 225)[48].

Como edições possíveis, temos a inserção, o apagamento, a substituição ou a reorganização de sequências de palavras. Como diferenças entre a TER e HTER, ressaltamos o fato de a primeira incluir um processo mais automático de comparação e cálculos de edição entre um texto traduzido por um sistema e uma tradução humana de referência via um algoritmo específico. A TER também só considera em suas edições correspondências exatas entre a tradução editada e a tradução de referência, havendo a versão TERp que passa a incorporar o reconhecimento de sinônimos e paráfrases no processo de edições. Já a HTER usa editores humanos e o cálculo de uma média de edições para que uma dada tradução gerada por um sistema de TM se assemelhe a uma tradução de referência, mantendo o significado e estando gramaticalmente correta na língua da tradução. Pede-se que os editores humanos editem minimamente a tradução gerada para que ela fique gramaticalmente correta, fluente na língua traduzida e apresente o mesmo sentido proposto pelo *gold standard*. Uma grande limitação do uso da HTER refere-se ao fato de ela ser dispendiosa e apresentar ruído ao necessitar do apoio humano por trás, ou seja, trata-se de uma métrica completamente dependente do apoio humano, podendo inclusive valer-se de uma subjetividade do tradutor humano. Portanto, para que seja passível de aplicação, é necessário que haja uma validação da tradução humana de referência ou se utilizem mais traduções humanas de referência, e que o número de editores envolvidos no processo da métrica seja de pelo menos 3 humanos para se amenizar o caráter subjetivo por

[48] "HTER is defined as the minimum number of edits needed to change a hypothesis so that it exactly matches one of the references, normalized by the average length of the references. Since we are concerned with the minimum number of edits needed to modify the hypothesis, we only measure the number of edits to the closest reference." (SNOVER *et al.*, 2006, p. 225).

trás do processo. Por outro lado, essa métrica gera resultados que se correlacionam melhor com julgamentos humanos do que a BLEU.

Há uma coleção de métricas da família MEANT que buscam trazer o aspecto semântico para a avaliação de TM. A métrica MEANT não é dispendiosa em seu processo e considera em sua constituição o conjunto de papéis semânticos e a "[...] estrutura de evento básica – quem fez o que para quem, quando, onde e por que" (Pradhan *et al.*, 2004, p. 1)[49]. Para os desenvolvedores da MEANT, uma boa tradução gerada por máquina deve ser capaz de capturar o sentido central da estrutura de evento instanciada no material textual a ser traduzido. A MEANT empenha-se em avaliar as correspondências entre os preenchedores dos papéis semânticos na tradução e compara-os com os do texto fonte e os das traduções de referência. São utilizados na MEANT os predicados, papéis e preenchedores semânticos no estilo *Propbank*. Lo e Wu (2011) descrevem como a MEANT funciona. Para os autores,

> Primeiro, a rotulação de função semântica é realizada (manual ou automaticamente) na tradução de referência e na tradução por máquina. As estruturas de conjuntos de papéis semânticos assim obtidas para a saída de TM são comparadas com as das traduções de referência, conjunto a conjunto, argumento por argumento. A acurácia da tradução via conjuntos de papéis é uma soma ponderada do número de argumentos traduzidos corretamente. Conceitualmente, a MEANT é definida em termos de *f-score*, com relação à *precision/recall* da acurácia da tradução de sentenças, calculada pela média da acurácia da tradução de todos os conjuntos na saída da TM através do número de conjuntos nas traduções por máquina e de referência. (LO; WU, 2011, p. 222)[50].

A partir dos testes realizados, Lo e Wu (2011) constataram que a MEANT foi eficiente na avaliação semântica da tradução por máquina em comparação com o texto de referência e indicou correlações com o julgamento humano semelhantes à HTER e se destacando de forma muito superior à BLEU.

Posteriormente, surge uma versão da MEANT também baseada em conjuntos de papéis argumentais, mas com uma abordagem entre línguas, a XMEANT. Essa variante da MEANT é multilíngue e, segundo Lo *et al.* (2014), ela pode ser obtida

[49] "[…] basic event structure – "who did what to whom, when, where and why". (Pradhan *et al*., 2004, p. 1).

[50] "First, semantic role labeling is performed (either manually or automatically) on both the reference translation and the machine translation. The semantic frame structures thus obtained for the MT output are compared to those in the reference translations, frame by frame, argument by argument. The frame translation accuracy is a weighted sum of the number of correctly translated arguments. Conceptually, MEANT is defined in terms of f-score, with respect to the precision/recall for sentence translation accuracy as calculated by averaging the translation accuracy for all frames in the MT output across the number of frames in the MT output/reference translations." (LO; WU, 2011, p. 222).

> [...] substituindo o modelo vetorial de contexto da MEANT monolíngue por probabilidades de tradução simples ao calcular semelhanças de preenchimento de papéis semânticos; e aprimorando a MEANT ainda mais ao incorporar restrições da Gramática Agrupada do Transdutor Inverso (BITG) para alinhar os *tokens* no preenchimento de papéis semânticos. (LO *et al.*, 2014, p. 765)[51].

A XMEANT utiliza *parsers* semânticos automáticos para a rotulação semântica, e realiza uma profunda integração semântica em um contexto multilíngue, ampliando o que havia sido proposto pela MEANT monolíngue, além de poder correlacionar-se melhor com os julgamentos humanos de adequação.

Outra métrica derivada da família MEANT é a MEANT 2.0 (LO, 2017). Trata-se de uma versão aprimorada da MEANT em que os pesos são atribuídos a cada palavra e atenta-se à ordem das palavras nas frases. Maiores pesos serão computados para as correspondências que se deram entre palavras de conteúdo e menores pesos para as correspondências entre palavras funcionais, dando importância à ordem das palavras no preenchimento de papéis semânticos na sentença traduzida como um todo. A MEANT 2.0 possui um custo de tempo reduzido pela metade em relação à MEANT, além de ser portátil, de código aberto e capaz de avaliar a qualidade de tradução para línguas pouco descritas computacionalmente de forma mais aprimorada que outras métricas como a BLEU.

No âmbito desta tese, serão aplicadas como métricas de avaliação de tradução dos sistemas aqui propostos e do sistema estado da arte a BLEU, TER e HTER, sendo esta última a que se mostra como a métrica mais adequada para detectar e avaliar tanto a correspondência semântica quanto a adequação sintática entre as traduções geradas por sistemas de TM em comparação com traduções *gold standard*. A escolha por submeter os sistemas à BLEU se deve ao reconhecimento dessa métrica e sua ampla utilização na avaliação de sistemas de TM. A não utilização de nenhuma das métricas da família MEANT que, embora considerem o aspecto semântico na avaliação, utilizam papéis semânticos do modelo específico *Propbank*, se deve ao fato de não haver uma modelagem disponível para o português nos moldes desse recurso com ampla cobertura, tornando-se inviável no momento para avaliar os sistemas aqui propostos. Por fim, utilizaremos a TER em comparação à HTER com o propósito de contrastar os métodos automático e humano no processo de edições envolvido na avaliação das métricas. No desenvolvimento de trabalhos futuros, pretende-se expandir o uso de métricas de avaliação e

[51] "[…] accomplished by replacing monolingual MEANT's context vector model with simple translation probabilities when computing similarities of semantic role fillers, and further improved by incorporating BITG constraints for aligning the tokens in semantic role fillers." (LO *et al.*, 2014, p. 765).

TM e considera-se o uso da METEOR (BANERJEE; LAVIE, 2005), chrF++ (POPOVIC, 2017) e novas métricas baseadas em vetores como BERTscore (SHIMANAKA *et. al.*, 2019).

A última das métricas que elencamos neste trabalho é a F-SEM, desenvolvida nos estudos desta tese e do programa *Google Summer of Code* 2019. Essa é uma métrica de avaliação de TM semântica que não requer o uso de traduções de referência ou correções humanas, buscando comparar diretamente o material textual traduzido com o material textual original. Para um teste inicial, ela se apoia em uma anotação manual de frames semânticos de um corpus paralelo em inglês, alemão e português (Legendas da *TED Talk Do schools kill creativity?*).

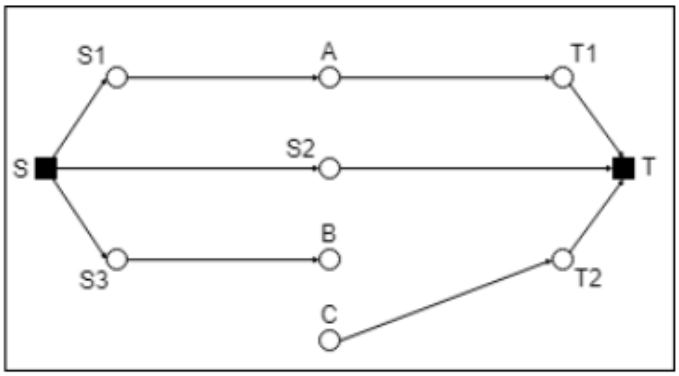
A F-SEM utiliza dois modelos teórico-metodológicos em sua implementação. O primeiro modelo é de tradução da primazia do frame (*Primacy of Frame*) (CZULO, 2017). Czulo (2017) propõe nesse modelo que a tradução e o original apresentam em alguma medida similaridades semânticas ao evocarem frames semelhantes ou próximos. O segundo modelo é o que utiliza a técnica da ativação espalhada (*Spreading Activation*). Czulo *et al.* (2019) descrevem que essa técnica é

> [...] um processo iterativo de propagação de energia com valor real de um ou mais nós de origem em uma rede usando links ponderados. Cada propagação é chamada de pulso. Basicamente, os pulsos são acionados a partir de um (ou mais) nó(s) inicial(is) e propagam-se através da rede, ativando os nós vinculados. Esse processo de ativação de mais e mais nós e verificação das condições de terminação é repetido pulso após pulso, até que todas as condições de terminação sejam atendidas, o que resulta em um estado final de ativação para a rede. (CZULO *et al.*, 2019, p. 30)[52].

Dados esses dois modelos, vejamos como a FrameNet pode ser ilustrada em termos de representação ou topologia para o funcionamento da F-SEM na Figura 18.

[52] "[…] an iterative process of propagating real-valued energy from one or more source nodes over a network using weighted links. Each propagation is called a pulse. Basically, pulses are triggered from one (or more) initial node(s) and propagates through the network, activating linked nodes. This process of activating more and more nodes and checking for termination conditions is repeated pulse after pulse, until all termination conditions are met, which results in a final activation state for the network." (CZULO *et al.*, 2019, p. 30).

Figura 18 – Topologia da FrameNet



Fonte: Designing a Frame-Semantic Machine Translation Evaluation Metric (CZULO *et al.*, 2019, p. 31).

Na Figura 18, a rede é composta da seguinte forma. O nó-fonte (S■) representa a sentença na língua-fonte. O nó-alvo (T■) refere-se à sentença na língua-alvo. Os nós S (○) são os frames evocados diretamente pela sentença-fonte (S1, S2 e S3). Os nós T (○) ilustram os frames evocados diretamente pela sentença-alvo (T1, T2 e T3). Por fim, os nós A(○), B(○) e C(○) simbolizam outros frames na hierarquia de frames. Partindo do funcionamento dessa rede, os autores apontam três situações possíveis. Na primeira situação, frames em comum são evocados pelas sentenças fonte e alvo (S2). Numa segunda situação, tanto a sentença original quanto a traduzida evocam frames que estão ligados em níveis superiores na rede (o frame S1 e o frame T1 estão ligados via relação frame-a-frame com o frame A). Uma última situação seria as sentenças fonte e alvo não compartilharem frames evocados, fazendo com que os frames T2 e C evocados pela sentença traduzida não sejam ativados.

Portanto, dada a insuficiência para lidar com questões semânticas das métricas de avaliação de TM anteriores, a F-SEM se debruça em sair do nível puramente lexical (como a BLEU), ampliando seu escopo para um nível semântico e funcional de avaliação. Ela também considera o fato de a tradução e o original não apresentarem uma correspondência de evocação de frames de um para um. Em sua versão final, a F-SEM busca ser totalmente automática ao realizar a anotação de frames das sentenças de forma automática (não requer envolvimento humano como a HTER). Ela captura o significado além da superfície utilizando frames da FrameNet, e não do estilo PropBank e sua hierarquia. Por fim, ela permite a avaliação de TM utilizando frames não apenas para os verbos principais das sentenças, mas para qualquer segmento anotado da sentença que evoca um frame. Entretanto, como a F-SEM foi inicialmente implementada para um corpus específico do projeto *Multilingual FrameNet*, não há a compatibilidade de uso para avaliar os sistemas de TM aqui propostos devido ao corpus

utilizado ter sido um corpus específico dos esportes apenas para duas línguas, sejam elas o inglês e o português. Sendo assim, utilizaremos as métricas BLEU, para avaliar formal e estruturalmente o desempenho dos sistemas de TM, e a TER/HTER, buscando uma avaliação mais adequada da correlação semântica e terminológica entre os textos traduzidos pelos sistemas de TM e o *gold standard*. Para estudos futuros e complementação desta tese, pretende-se implementar a F-SEM para avaliação de sistemas de TM devido à consideração dada aos aspectos semânticos da tradução.

Passamos agora ao Capítulo 3, em que abordaremos as teorias linguísticas que norteiam este trabalho, sejam elas a Semântica de Frames e as Relações Qualia.

3. MODELAGEM LINGUÍSTICO-COMPUTACIONAL DO LÉXICO

A Linguística Cognitiva se estabelece enquanto área de pesquisa na década de 1980. Fillmore, Lakoff, Langacker, Talmy, entre outros, são autores de destaque para essa área. Para Geeraerts e Cuyckens (2007), esses autores "[...] compreendem a língua como um instrumento de organização, processamento e transmissão de informações" (GEERAERTS; CUYCKENS, 2007, p. 3)[53]. A Linguística Cognitiva se posiciona como uma área que propõe uma relação direta entre a linguagem e a cognição, passando pela experiência cultural, social e individual, e por seu reflexo direto no mundo. Vejamos a Semântica de Frames, subárea da Linguística Cognitiva no campo da significação e foco desta tese.

3.1 SEMÂNTICA DE FRAMES E SUAS APLICAÇÕES TECNOLÓGICAS

Como uma abordagem empírica para o estudo do significado, a Semântica de Frames passa a existir considerando as experiências categorizadas pelos falantes através das escolhas que fazem ao usar a língua. Para Fillmore (1982), a Semântica de Frames

> [...] oferece um modo específico de olhar para os significados das palavras, bem como um meio de caracterizar os princípios para a criação de novas palavras ou sintagmas, para acrescentar novos significados às palavras, e para a composição dos significados dos elementos de um texto no significado total do texto. (FILLMORE, 1982, p. 111)[54].

Ao mencionar o significado total do texto, constata-se a noção de contexto ou cena contextual, o que envolve também as informações não linguísticas relacionadas aos frames ativados. Para a Semântica de Frames, "[...] as formas linguísticas evocam ou ativam o conhecimento de frames, e como os frames ativados podem ser integrados em uma compreensão das passagens que contêm essas formas." (FILLMORE; BAKER, 2010, p. 317)[55].

A partir daí, emerge o conceito basilar para a Semântica de Frames, que é o *frame*. Esse conceito ganha destaque a partir de estudos realizados nas décadas de 1970 e 1980, perpassando diversas áreas de conhecimento, tais como a psicologia, a sociologia, a inteligência artificial e

---

[53] "[…] they see language as an instrument for organizing, processing, and conveying information." (GEERAERTS; CUYCKENS, 2007, p. 3).

[54] "[…] offers a particular way of looking at word meanings, as well as a way of characterizing principles for creating new words and phrases, for adding new meaning to words, and for assembling the meaning of elements in a text into the total meaning of the text." (FILLMORE, 1982, p. 111).

[55] "[…] linguistic forms evoke or activate frame knowledge, and how the frames thus activated can be integrated into an understanding of the passages that contain these forms." (FILLMORE; BAKER, 2010, p. 317).

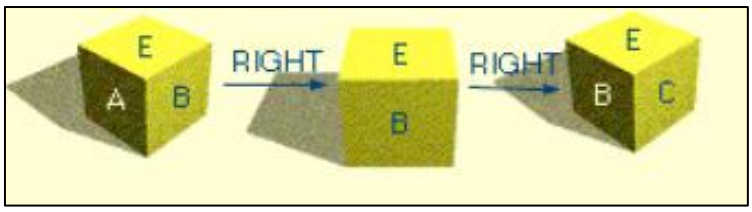
a linguística. Na área de Inteligência Artificial (A.I.), Minsky (1974) aplica o conceito de frames como estruturas de informação que estão implícitas e representam uma situação estereotipada. O autor postula que

> Um frame é uma estrutura de dados utilizada para representar uma situação estereotipada, como estar em um certo tipo de sala de estar, ou ir à festa de aniversário de uma criança. Anexados a cada frame estão vários tipos de informações. Algumas dessas informações são sobre como usar o frame. Algumas são sobre o que se pode esperar que aconteça a seguir. Algumas são sobre o que fazer se essas expectativas não são confirmadas. (MINSKY, 1974, p. 1)[56].

Com base nas afirmações do autor, ao se deparar com uma situação, deve-se fazer uma mudança substancial na visão do problema e selecionar da memória a cena que mais se adeque ao contexto em que se encontra. Para o autor, o frame está ligado à experiência e ao pensamento esquemático, em que situações mais complicadas são ligadas a estruturas de frame estereotipadas.

Minsky compara o sistema de frames com a representação das diferentes faces ou perspectivas de um mesmo cubo. Ao analisarmos a Figura 19, proposta pelo autor, perceberemos um mesmo cubo representado três vezes com uma sutil mudança de posição.

Figura 19 – Representação do sistema de frames na forma de um cubo e suas diferentes faces e perspectivas



Fonte: A Framework for Representing Knowledge. (MINSKY, 1974, p.8).

Na Figura 19, Minsky propõe que, ao começarmos a observar o cubo mais à esquerda e formos caminhando para a direita, notaremos diferentes perspectivas e faces. Quando estamos mais à esquerda, vemos claramente as faces A, B e E. Mas ao caminharmos para a direita, a face A "desaparece" e passamos a ver apenas as faces B, C e E. Segundo o autor, durante esse processo, temos que reanalisar a imagem, (i) perdendo o conhecimento de "A", (ii)

[56] "A frame is a data-structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed." (MINSKY, 1974, p. 1).

recomputando “B”, e (iii) computando a descrição de “C”. Essa representação nos levaria a uma visão mais ampla do sistema de frames em que cada frame seria representado por uma perspectiva sobre o cubo. Para o autor, “[...] uma grande coleção de sistemas de frames é armazenada na memória permanente, e um deles é evocado quando a evidência e a expectativa tornam plausível que ele se encaixe na cena em destaque.” (MINSKY, 1974, p. 8)[57]. A noção de perspectiva será melhor desenvolvida mais adiante na descrição da FrameNet.

De forma análoga, Fillmore (1975; 1977; 1982; 1985) defende que um frame é um sistema de categorias estruturadas conforme um contexto que as motive. Para o autor, certas palavras existem para dar aos participantes da cena comunicativa acesso ao conhecimento. Posteriormente, essa concepção contribui para a construção do conceito de evocação. Um falante aplica um frame a uma situação e demonstra sua intenção de uso de tal frame ao selecionar palavras que evocam esse frame. A noção de frame proposta por Fillmore (1982) ocupa-se de um termo mais abrangente e genérico que cobre uma série de conceitos amplamente conhecidos na literatura de Compreensão de Língua Natural (*Natural Language Understanding* – NLU), tais como esquema, *script*, cenário, modelo cognitivo idealizado, entre outros. Como já apontado na Introdução desta tese, o autor postula o conceito de frame como sendo

> [...] qualquer sistema de conceitos relacionados de tal modo que, para entender qualquer um deles, é preciso entender toda a estrutura na qual se enquadram; quando um dos elementos dessa estrutura é introduzido em um texto, todos os outros elementos serão disponibilizados automaticamente. (FILLMORE, 1982, p. 111)[58].

O linguista trabalha com o conceito de frame relacionando os elementos participantes de cada cena entre si, mesmo que nem todos apareçam instanciados na sentença. Ao mencionar que é necessário entender toda a estrutura na qual os elementos se encontram presentes, o autor refere-se aos contextos linguístico e não-linguístico, o que para Minsky poderia ser a situação estereotipada.

Fillmore inicialmente separa os frames em duas categorias: os frames cognitivos e os linguísticos. Os frames cognitivos “[...] se referem a estruturas que pessoas invocam para dar sentido a suas observações” (FILLMORE, 2008, p. 2)[59], enquanto os frames linguísticos “[...]

---

[57] “[…] a great collection of frame systems is stored in permanent memory, and one of them is evoked when evidence and expectation make it plausible that the scene in view will fit it.” (MINSKY, 1974, p. 8).

[58] “[...] any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, all of the others are automatically made available”. (FILLMORE, 1982, p. 111).

[59] “[…] to refer to the structures that people invoke to make sense of their observations.” (FILLMORE, 2008, p. 2).

se referem àqueles frames que estão convencionalmente associados a um material linguístico específico. As pessoas invocam os frames e as formas linguísticas evocam os frames nas mentes dos falantes de uma língua" (FILLMORE, 2008, p. 2)[60]. Com o passar do tempo, percebe-se uma fusão entre os conceitos de frame com quaisquer tipos de conhecimento que podem ser evocados por meios lexicais ou linguísticos, ou invocados pelo falante, havendo uma conexão entre eles.

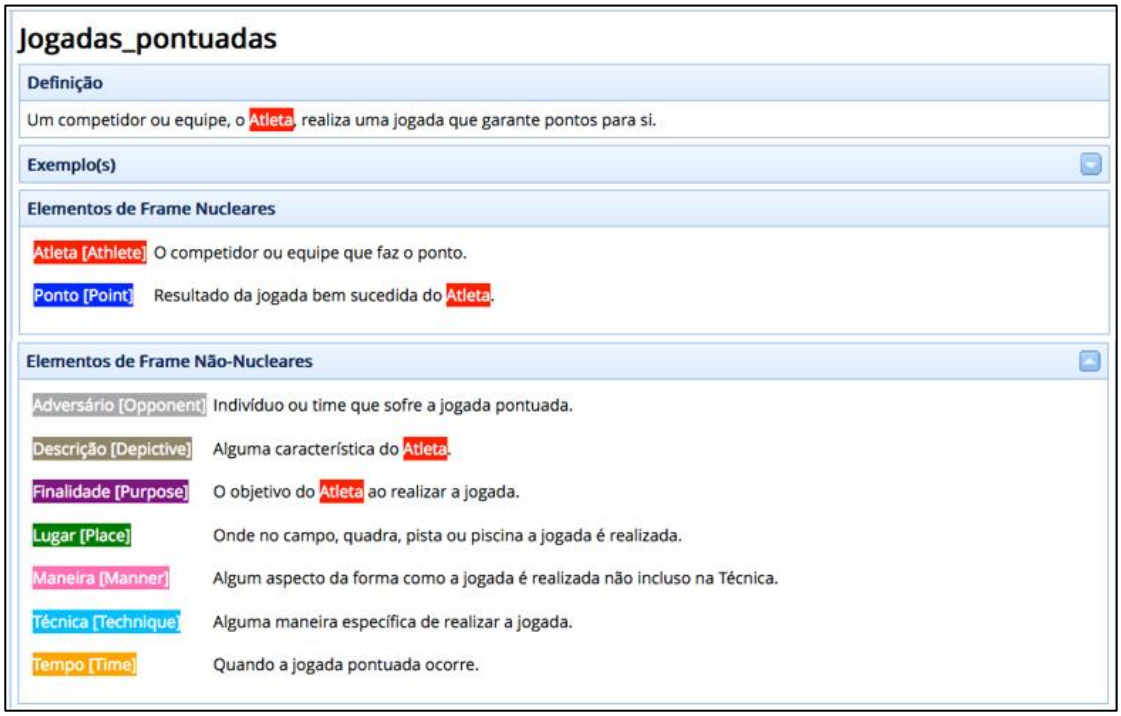
A partir de 1997, a teoria da Semântica de Frames passou a ser aplicada a uma iniciativa de descrição lexicográfica da língua inglesa denominada FrameNet. Tal projeto, desenvolvido no ICSI (*International Computer Science Institute*), em Berkeley, Califórnia, é a aplicação mais proeminente da Semântica de Frames. Constitui-se de uma base computacional de dados lexicográficos fundada em anotação de sentenças e textos em corpora, que vem sendo expandida em projetos paralelos de construção de *framenets* para diversas línguas pelo mundo, como o espanhol, o alemão, o chinês, o japonês, o sueco e o português do Brasil. Exploremos as características da base de dados do projeto FrameNet Brasil. Na Figura 20, temos a representação de um frame na *webtool 3.0* da FrameNet Brasil[61].

A Figura 20 ilustra o frame `Jogadas_pontuadas`. Logo abaixo do nome do frame, há uma definição composta de uma descrição textual do que ele representa. Os participantes correspondem aos Elementos de Frame (EFs) acompanhados de suas definições. Como EF, entendemos qualquer papel semântico definido especificamente no frame e que fornece uma informação adicional à estrutura da sentença. Eles podem ser divididos em Nucleares (*Core*) e Não-nucleares (*Non-core*). Os EFs nucleares são aqueles considerados essenciais na definição de um frame para que ele possa existir cognitivamente. Eles tornam um frame único e diferente dos demais. Já os Não-nucleares são EFs que enriquecem o frame com detalhes adicionais. Eles são de dois tipos: os Periféricos (*Peripheral*) e os Extra-temáticos (*Extra thematic*). "Os EFs periféricos são aqueles que não introduzem eventos extras, independentes ou diferentes do evento principal" (RUPPENHOFER *et al.*, 2006, p. 24)[62]. Eles podem aparecer em diversos frames caracterizando ideias de tempo, meio, lugar, maneira, grau, entre outras. Já os EFs extra-temáticos "[...] situam um evento em um cenário de outro estado de coisas, seja de um evento

[60] "[…] to refer to those frames that are conventionally associated with specific linguistic material. People invoke frames, and that linguistic forms evoke frames in the minds of those who know the language". (FILLMORE, 2008, p. 2).
[61] As ilustrações apresentadas neste trabalho que envolvem frames, EFs e ULs são extraídas especificamente da Webtool da FrameNet Brasil.
[62] "Frame elements that do not introduce additional, independent or distinct events from the main reported event are characterized as peripheral." (RUPPENHOFER *et al.*, 2006, p. 24).

real ou do mesmo tipo, conforme ilustrado em **Iteração**, ou evocando um frame maior no qual o estado de coisas relatado está incorporado, como mostrado em **Evento_continente**." (RUPPENHOFER *et al.*, 2006, p. 24)[63]. Trata-se de EFs que são introduzidos por certos elementos construcionais ou evocam frames separadamente. Dentro do frame `Jogadas_pontuadas`, ilustrado na Figura 20, temos dois EFs nucleares que são Atleta e Ponto. Eles são essenciais na constituição do frame. Já os EFs periféricos desse frame são Adversário, Finalidade, Lugar, Maneira, Técnica e Tempo. Esse frame também apresenta o EF extra-temático Depictivo.

Figura 20 – Representação do frame `Jogadas_pontuadas`



Fonte: Webtool da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

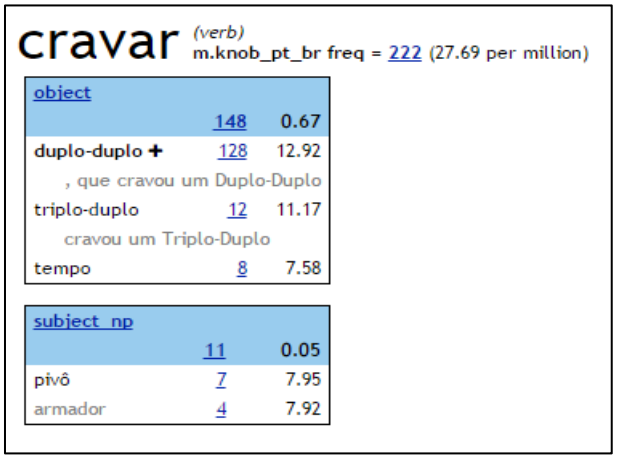
As Unidades Lexicais (ULs), que são pareamentos entre lemas e frames, possuem cada uma um significado específico. Elas são os elementos evocadores de frames e modeladas dentro do frame que evocam. Exemplos de ULs que evocam o frame `Jogadas_pontuadas` incluem abertura.n, abrir marcador.v, abrir placar.v, ace.n e acertar.v.

Na FrameNet, há anotação sintático-semântica de material textual em corpora, o que marca o comportamento linguístico das ULs em contexto. Essa anotação pode ocorrer a partir

[63] "[...] situate an event against a backdrop of another state of affairs, either of an actual event or state of the same type, as illustrated with **Iteration**, or by evoking a larger frame within which the reported state of affairs is embedded, as shown for **Containing_event**." (RUPPENHOFER *et al.*, 2006, p. 24).

de duas metodologias: a lexicográfica e a de texto corrido. Na anotação lexicográfica, a partir de corpora previamente selecionados, extraem-se sentenças candidatas à anotação de uma UL alvo. A ferramenta utilizada na extração das sentenças é o *WordSketch* presente no *SketchEngine* (KILGARRIFF *et al*., 2004). Essa ferramenta não apenas busca pelo lema, ou seja, se for um verbo, trata-se de todas as formas flexionadas ou não desse verbo, mas também apresenta os dados separados por configuração sintática. O princípio da anotação lexicográfica é o de que o anotador pode ter uma intuição sobre o funcionamento dessa UL com base em sua configuração sintática. A Figura 21 traz a busca por uma palavra específica dos esportes em corpora e o funcionamento da ferramenta *WordSketch* do *SketchEngine*.

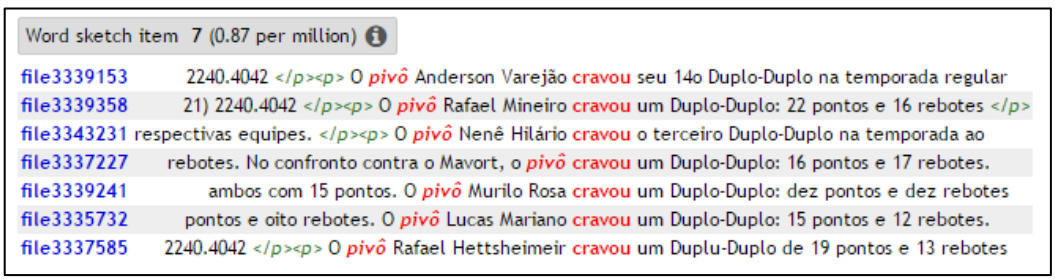
Figura 21 – UL cravar.v na ferramenta WordSketch



Fonte: Ferramenta *WordSketch* do SketchEngine. (https://app.sketchengine.eu/).

Na Figura 21, atestamos que, ao se pesquisar pela UL **cravar.v** no corpus, há um total de 222 ocorrências. Constata-se que os objetos mais frequentes que ocorrem com essa UL são "duplo-duplo", "triplo-duplo" e "tempo". Como sujeitos mais frequentes, temos as posições "pivô" e "armador" relacionadas ao basquete e a outros esportes. Ao clicarmos em "pivô" como sujeito de "cravar", temos os resultados apresentados na Figura 22.

Figura 22 – Sentenças que instanciam a UL cravar.v a partir do sujeito "pivô"



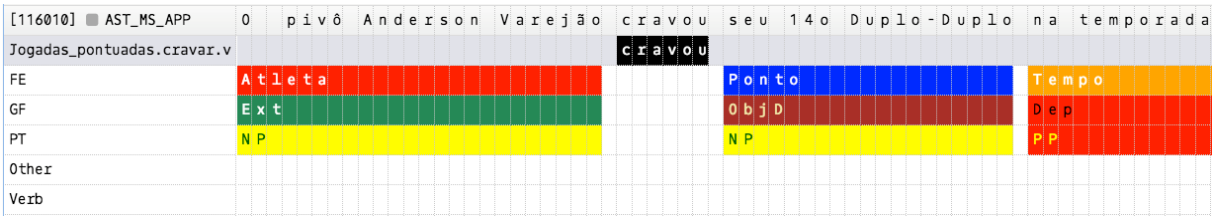
Fonte: Ferramenta de Uso de Corpora *SketchEngine*. (https://app.sketchengine.eu/).

O anotador, analisando os dados obtidos na Figura 22, notaria que há uma certa configuração sintática que poderia lhe oferecer pistas do comportamento sintático-semântico da UL **cravar.v** no frame `Jogadas_pontuadas`. Existe um Externo e Atleta representados pela palavra "pivô" e o Objeto Direto e Pontuação marcados por "Duplo-Duplo", número de "pontos" e "rebotes". A partir dessa demonstração do WordSketch, acredita-se que a ferramenta contribui para se minimizar o consumo de tempo na seleção de sentenças de uma amostra.

Outro tipo de anotação é a de texto corrido. Seleciona-se um texto que vai ser anotado e todas as ULs que nele se manifestarem serão anotadas. Assim, uma mesma sentença pode resultar em diversas anotações, cada uma tendo uma UL distinta como alvo.

Em ambos os casos, o processo de anotação é feito através da WebTool da FrameNet Brasil. Há uma interface gráfica que permite ao anotador atribuir rótulos coloridos que representam os EFs do frame evocado pela UL alvo, bem como a função gramatical (GF) e o tipo sintagmático (PT) do material linguístico que instancia os EFs. A Figura 23 ilustra a anotação lexicográfica na Webtool da FrameNet Brasil.

Figura 23 – Anotação lexicográfica de uma sentença na Webtool



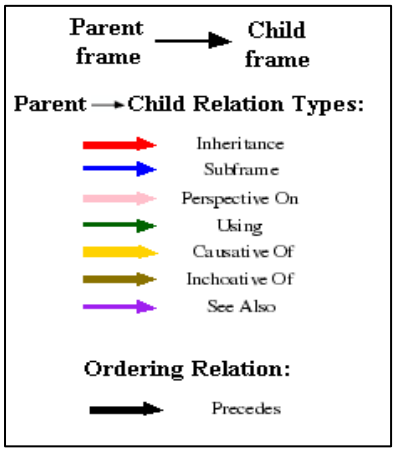
Fonte: Webtool da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

Na sentença exemplificada na Figura 23, a UL **cravar.v** evoca o frame `Jogadas_pontuadas`. A anotação é feita, minimamente, em três camadas. A primeira camada marca o EF correspondente à porção de texto que se relaciona à UL alvo da anotação. No exemplo em destaque, o fragmento do texto "O pivô Anderson Varejão" é marcado com o EF Atleta. A segunda camada é a Função Gramatical (GF). Na amostra em questão, "O pivô Anderson Varejão" seria o Argumento Externo da UL **cravar.v**. Já a terceira camada representa o Tipo Sintagmático (PT), que nesse segmento seria um Sintagma Nominal (NP). Continuando a anotação da sentença, temos que a porção textual "seu 14º Duplo-Duplo" é marcada pelo EF Ponto, pela GF Objeto Direto e pelo PT Sintagma Nominal. Já o último fragmento de texto da

sentença representado por “na temporada” é anotado como EF Tempo, GF Dependente e PT Sintagma Preposicionado. Há ainda uma camada “Outro” e uma específica “Verbo”. No caso da anotação de uma UL verbal, essas camadas não são obrigatórias, mas denotam características especiais dos segmentos como a anotação de pronomes relativos e seus antecedentes. A partir dessa anotação em camadas, extraímos padrões de valência sintático-semânticos do comportamento das unidades lexicais em contexto, dado o fato de que as sentenças anotadas partem de um corpus compilado de dados reais da língua.

A FrameNet, enquanto uma rede semântica, conta também com relações entre frames, sejam elas: Herança, Perspectiva, Uso, Subframe, Precedência, Causativo_de, Incoativo_de e a metarrelação Veja_também. Analisemos os tipos de relações frame-a-frame a partir da Figura 24.

Figura 24 – Legenda de relações frame-a-frame da FrameNet



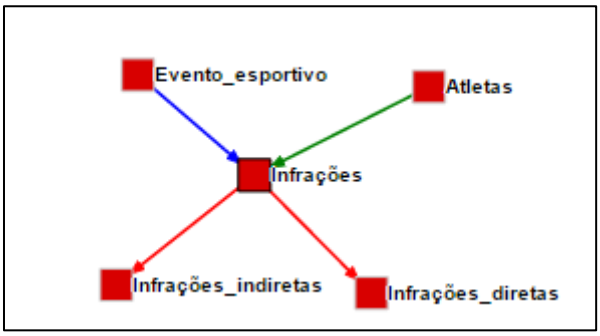
Fonte: Berkeley FrameNet. (https://framenet2.icsi.berkeley.edu/FrameGrapher/LegendPopup.html).

A Figura 24 denota que cada relação é modelada na FrameNet seguindo um padrão específico de cores. A relação de herança (*Inheritance*), representada pela seta vermelha, ocorre quando um frame filho elabora um ou mais elementos do frame pai. Diz-se que o frame filho nessa relação é um “tipo” do frame pai. A relação de Perspectiva (*Perspective_on*), ilustrada pela seta rosa, indica a presença de pelo menos dois pontos de vista diferentes de um mesmo frame neutro. A relação de Uso (*Using*), marcada pela seta verde, propõe o fato do um frame filho pressupor um ou mais frames pais como pano de fundo. Para frames mais complexos, temos a relação de Subframe (*Subframe*), sinalizada pela seta azul, que fragmenta o frame mais complexo em frames separados. A relação de Precedência (*Precedes*), indicada pela seta preta, ocorre quando estamos lidando com frames que estão dispostos numa sequência de estados ou

eventos que contribuem para a definição de um certo estado de coisas. Ainda temos as relações Causativo_de (*Causative_of*), marcada pela seta amarela, e Incoativo_de (*Inchoative_of*), ilustrada pela seta marrom, que exprimem relações de causa e mudança de estado respectivamente. Por último, no caso de frames muito semelhantes entre si, propõe-se a relação Veja_também (*See_also*), marcada pela seta roxa, de modo a auxiliar o usuário na compreensão das especificidades de cada frame ao indicar outros frames para análise.

Na base de dados da FrameNet, há uma representação gráfica da rede de frames e suas relações denominada Grapher. A partir do Grapher, há a geração de gráficos que ilustram desde a relação entre dois frames, até as relações entre todos os frames modelados na base de dados. Na seção da 4.2, Estrutura da Base de Conhecimento, será apresentada e discutida toda a modelagem de frames e relações conduzida no âmbito desta tese para os frames do domínio dos Esportes. Notemos na Figura 25 um exemplo de gráfico de frames e relações gerado pela ferramenta Grapher da FrameNet Brasil.

Figura 25 – Relações entre frames na FrameNet Brasil



Fonte: Ferramenta Grapher da Webtool da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

A Figura 25 traz a representação das relações modeladas entre alguns frames no domínio dos Esportes extraídas do Grapher da FrameNet Brasil. Constata-se que o frame `Infrações` é um subframe (➔) do frame `Evento_esportivo`, ou seja, dentro da grande cena de evento esportivo, temos eventos menores e as infrações representam um desses eventos. O frame `Atletas` possui uma relação de Uso (➔) com o frame `Infrações`, uma vez que o frame filho `Infrações` pressupõe em parte dele a presença de EFs e informações do frame pai `Atletas`. Portanto, estabelece-se aí a relação de Uso entre eles. Por fim, temos a relação de herança (➔) ilustrada entre o frame pai `Infrações` e os frames filhos `Infrações_indiretas` e `Infrações_diretas`. É importante ressaltar que, na relação

de herança, os frames filhos elaboram características do frame pai, portanto especificam EFs do frame pai.

Abordamos nessa seção a Semântica de Frames, o conceito de frame (EFs e ULs), a FrameNet e suas especificidades (anotação, relações entre frames e relações *etc*.) e a busca por sentenças em corpora (SketchEngine). Ainda para o propósito de análise deste trabalho, se faz necessária a implementação de outras relações fundamentais que estabelecem uma ponte semântica entre as ULs, quais sejam as relações qualia. Na seção 3.2 abordaremos essas relações através da Teoria do Léxico Gerativo.

3.2 TEORIA DO LÉXICO GERATIVO

A Teoria do Léxico Gerativo (TLG) (PUSTEJOVSKY, 1995) surge como uma abordagem que lida com a semântica das palavras, como elas se combinam, o que denotam, além de mecanismos peculiares como a polissemia e a coerção de tipos. O avanço da teoria se deve a uma insatisfação de muitos linguistas teóricos e computacionais com a caracterização do léxico como um conjunto fechado e estático de traços sintáticos, morfológicos e semânticos.

Pustejovsky (1995) apresenta a TLG como "[...] uma abordagem decomposicional em que os itens lexicais são minimamente decompostos em formas estruturadas, em vez de um conjunto de traços." (PUSTEJOVSKY, 1995, p. 58)[64]. O aspecto gerativo da TLG diz respeito ao fato de a teoria atribuir uma descrição estrutural ao léxico através de níveis de descrição semântica. Sendo uma abordagem decomposicional do significado, a TLG lida com o aspecto criativo de combinações lexicais e a representação semântica a partir da descrição de uma série de componentes primitivos do significado. Partindo de uma modulação e um refinamento em nível pragmático, essa abordagem sofre fortes influências de fatores situacionais e contextuais. Antes de detalharmos os níveis de representação semântica propostos pela TLG, abordaremos o conceito de entidade, fundamental para este trabalho, visto que trabalhamos com melhorias em sistemas de tradução com ênfase nas entidades.

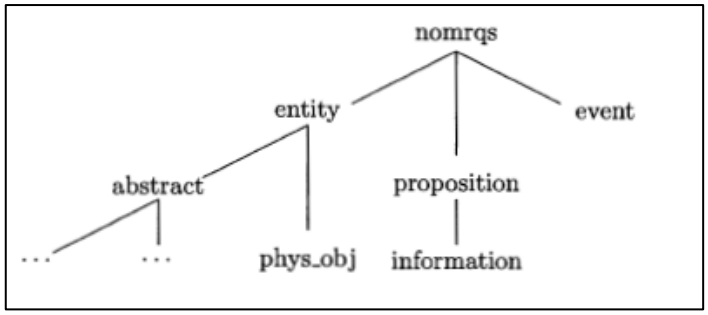
3.2.1 Entidades e Entidade Nomeada (EN)

[64] "[…] an approach to decomposition, where lexical items are minimally decomposed into structured forms rather than a set of features." (PUSTEJOVSKY, 1995, p. 58).

Entidade pode ser compreendida como uma categoria que representa coisas comuns como objetos, pessoas e lugares, também podendo ser nomeada. Pustejovsky e Stubbs (2012) apresentam uma Entidade Nomeada (EN) como "[...] uma entidade (um objeto no mundo) que tem um nome que o identifica exclusivamente pelo nome, por um apelido, abreviatura e assim por diante. 'O'Reilly', 'Universidade Brandeis', 'Mount Hood', 'IBM' e 'Vice-Presidente' são exemplos de ENs." (PUSTEJOVSKY; STUBBS, 2012, p. 71)[65]. Para este trabalho, interessam as entidades comuns, ou seja, os substantivos comuns que indicam objetos, pessoas, lugares etc., os quais não devem ser confundidos com ENs, que se manifestam através de substantivos próprios.

Pustejovsky (1995) propõe que os nomes de entidade podem ser representados em termos de traços através de estruturas de representação semântica. Vejamos na Figura 26 esse esquema de representação semântica de uma hierarquia de tipos.

Figura 26 – Fragmento de uma hierarquia de tipos



Fonte: The Generative Lexicon. (PUSTEJOVSKY, 1995, p.90).

A Figura 26 ilustra uma estrutura *lattice*, em que *nomrqs* representa um elo superior entre os mais diversos tipos de entidades (*entity*), proposições (*proposition*) e eventos (*event*). As entidades podem ser caracterizadas por diversos tipos de traços. No exemplo em questão, temos apenas uma divisão entre uma entidade abstrata e a entidade enquanto um objeto físico.

Helbig (2006) expande o conceito de entidade ao propor uma classificação de entidades em termos de uma ontologia conceitual. Ao estabelecer um sistema de representação do conhecimento, o autor diz estar tratando de unidades representativas de conceitos (tipos) e não necessariamente de objetos reais no mundo. As Entidades (ent) aparecem no topo da ontologia

[65] "[…] an entity (an object in the world) that has a name which uniquely identifies it by name, nickname, abbreviation, and so on. 'O'Reilly', 'Brandeis University', 'Mount Hood', 'IBM', and 'Vice President' are all examples of NEs." (PUSTEJOVSKY; STUBBS, 2012, p. 71).

como o tipo mais genérico de entidades conceituais, compreendendo todas as coisas que podem ser afirmadas. Elas são representadas por nós na Rede Semântica ilustrada na Figura 27.

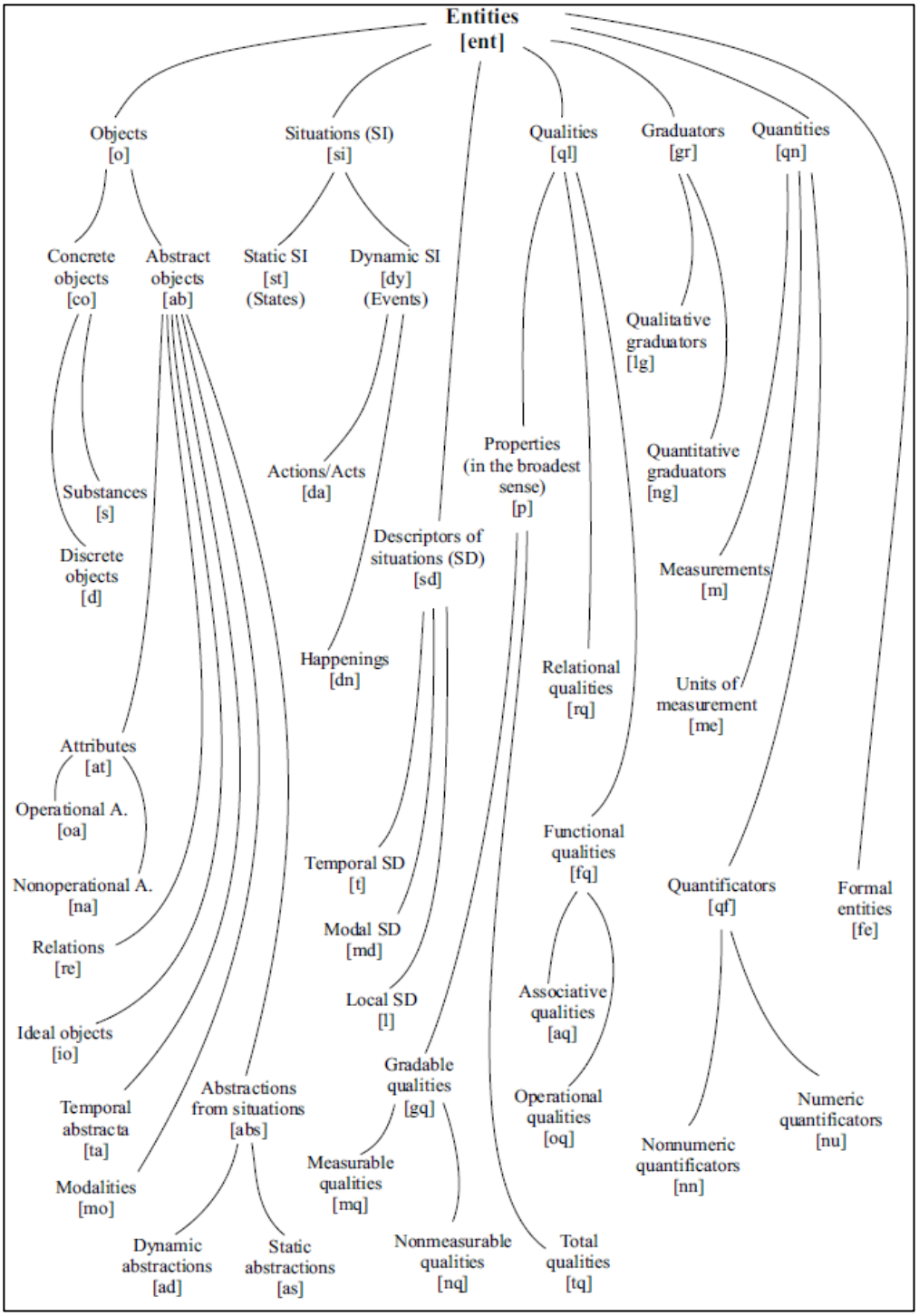
A Figura 27 representa um esquema proposto por Helbig (2006) e sua descrição se dará da direita para esquerda e de cima para baixo, no intuito de que a descrição dos objetos fique para o final e possamos discutir acerca de sua importância no âmbito desta pesquisa. A análise da Figura 27 também se dará com as definições do autor, com seus exemplos e alguns outros próprios deste trabalho.

No topo da hierarquia, há sete tipos principais de entidades, respectivamente: (i) Entidades Formais (fe), (ii) Quantidades (qn), (iii) Elementos de Gradação (gr), (iv) Qualidades (ql), (v) Descritores Situacionais (sd), (vi) Situações ou Estados de Coisas (si) e (vii) Objetos (o). Vejamos com mais detalhes as propriedades de cada um dos subtipos de entidades:

i. As Entidades Formais são utilizadas para designar objetos extralinguísticos tais como gráficos, fórmulas, tabelas, ilustrações, dentre outros tipos de estruturas que desempenham um papel importante em certos documentos. Essas entidades normalmente estão ligadas a expressões textuais referenciais Ex.: “ao fundo da imagem”, “as barras mais à esquerda do gráfico”.
ii. As Quantidades são entidades que expressam os diferentes aspectos de um conceito em termos de Quantificadores (qf), Unidades de Medida (me) e Medidas (m). Os quantificadores se referem a quantidades numéricas (nu), ex.: “trinta e dois”, e não-numéricas (nn), ex.: “algumas”. As unidades de medida, ex.: litro, m, KW, são usadas com números e quantidades para a especificação de medidas, ex.: 3kg.
iii. Os Elementos de Gradação marcam o grau ou a intensidade das propriedades e das quantidades. Os Elementos de Gradação Qualitativos (lg) lidam com a descrição das propriedades, ex.: bastante, e os Elementos de Gradação Quantitativos (ng) se ligam às quantidades, ex.: cerca de.
iv. As Qualidades são as entidades que caracterizam as propriedades dos objetos e das situações e podem ser interpretadas como sendo:
    - Propriedades em sentido estrito (p), com qualidades totais (tq), ex.: morto, e qualidades gradativas (gq), ex.: bonito. Dentre as qualidades gradativas estão as propriedades mensuráveis (mg), ex.: alto, e as propriedades não mensuráveis, mas passíveis de julgamento (nq), ex.: invejoso.
    - Qualidades Relacionais (rq): são as propriedades que estabelecem uma relação entre dois elementos. Ex.: “inverso”.

- Qualidades Funcionais (fq): tem seu significado pleno apenas se ligadas a entidades. Elas podem ser Propriedades Semanticamente Associativas (aq), ex.: propriedades químicas, e Propriedades Operacionais (oq), que descrevem posições operacionalmente, ex.: “sétimo” e “penúltimo”.

Figura 27 – Ontologia de Tipos com foco nas Entidades



Fonte: Knowledge Representation and the Semantics of Natural Language. (HELBIG, 2006, p.410).

v. Os Descritores Situacionais validam as situações em termos de sua incorporação espaço-temporal. Eles se apresentam em três tipos: tempo da situação (t), local da situação (l) e especificações modais (md). Os descritores temporais envolvem mudanças dinâmicas no mundo com sequências de eventos, ex.: “às quartas”, “em 2007”. Os descritores locais englobam locais conectados a objetos ou que constituem situações em sentido mais amplo, ex.: “no quintal”. Por último, as modalidades compreendem conceitos que expressam a posição do orador ou a opinião comum em relação à validade dos estados de coisas ou situações, ex.: “provavelmente”, “necessário”.

vi. As Situações ou Estados de Coisas espelham uma coleção de objetos, seus modos de ser e mudanças que sofrem. Podem ser divididas em Situações Estáticas (estados) e Dinâmicas (eventos). As Situações Estáticas (st) envolvem os estados físicos e psíquicos expressos por substantivos como “doença” em “ter uma doença grave”, verbos de estado em particípios passivos como “estar doente”, ou em construções predicativas como “está quente”. Por outro lado, as Situações Dinâmicas (dy), também denominadas Eventos, são divididas em Ações (da) ou Acontecimentos (dn). As Ações são realizadas ativamente por um agente, ex.: “cantar”. Os Acontecimentos têm causas, mas não há um agente associado ativamente ao evento, ex.: chuva.

vii. Os Objetos se dividem em dois tipos: os Objetos Concretos (co) podem ser percebidos através dos sentidos e os Objetos Abstratos (ab) não podem ser notados sensorialmente. Os Objetos Concretos podem se dividir em Substâncias (s), com uma extensão semi-contínua, divisíveis e não contáveis, ex.: “água”, e os Objetos Discretos (d), contáveis, mas não divisíveis como uma “casa”. Já os Objetos Abstratos são produtos do raciocínio humano e se distinguem entre: Objetos Situacionais (abs), Atributos (at), Relações (re), Objetos Ideais (io), Objetos Temporais Abstratos (ta) e Modalidades (mo). Os Objetos Situacionais representam situações elevadas por abstração ao status cognitivo dos objetos e podem ser divididos em dois tipos: as Abstrações de Situações Dinâmicas (ad), como corrida, roubo, movimento, e as Abstrações de Situações Estáticas (as), como calma e equilíbrio. Os Atributos podem ser divididos em Mensuráveis Operacionalmente (oa), como a altura e o peso, e os Não-Mensuráveis Operacionalmente (na), como a forma e a flexibilidade. Já as Relações podem ser exemplificadas por causalidade,

sinonímia, similaridade etc. Dentro dos Objetos Ideais, aloquemos entidades como religião, justiça, categoria etc. Para os Objetos Temporais Abstratos, citemos a Renascença e a Páscoa. Por último, como Modalidades, temos entidades como probabilidade e permissão.

Vimos, na Figura 27 e nas explicações propostas por Helbig (2006), o detalhamento do conceito de entidade numa ontologia conceitual de representação do conhecimento. Ao lidarmos com Entidade no escopo deste trabalho, traça-se uma analogia ao que o autor propõe como Objeto. Mais adiante, estabeleceremos uma comparação entre alguns elementos da Figura 27 com algumas das estruturas que utilizaremos neste trabalho. Passemos agora aos níveis de representação semântica da Teoria do Léxico Gerativo.

3.2.2 Níveis de Representação Semântica da TLG

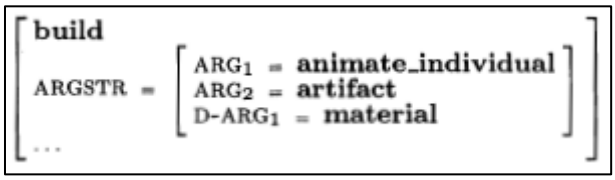
Dentro da TLG, Pustejovsky (1995) postula que o léxico gerativo pode ser descrito como um sistema que abrange pelo menos quatro níveis de representação semântica.

> Esses incluem a noção de estrutura argumental, que especifica o número e tipo de argumentos que um item lexical porta; uma estrutura de evento de riqueza suficiente para caracterizar não apenas o tipo de evento básico de um item lexical, mas também uma estrutura subeventual e interna; uma estrutura qualia, representando os diferentes modos de predicação possíveis de um item lexical; e, uma estrutura de herança lexical, que identifica como uma estrutura lexical está relacionada com outras estruturas em um dicionário, apesar de ela ser construída. (PUSTEJOVSKY, 1995, p. 58)[66].

Detalhemos os níveis de representação semântica propostos. A estrutura argumental, além de especificar o número e os tipos de argumentos lógicos, expressa como esses argumentos são realizados sintaticamente. Essa manifestação sintática dos argumentos faz com que a estrutura argumental seja o nível de representação de mais fácil compreensão. Vejamos uma representação de estrutura argumental.

[66] "These include the notion of argument structure, which specifies the number and type of arguments that a lexical item carries; an event structure of sufficient richness to characterize not only the basic event type of a lexical item, but also internal, subeventual structure; a qualia structure, representing the different modes of predication possible with a lexical item; and, a lexical inheritance structure, which identifies how a lexical structure is related to other structures in the dictionary, however it is constructed". (PUSTEJOVSKY, 1995, p. 58).

Figura 28 – Representação da estrutura argumental da palavra *build* (construir)

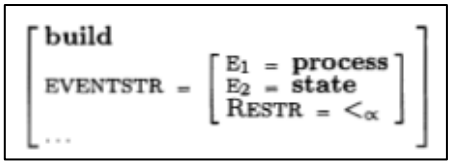


Fonte: The Generative Lexicon. (PUSTEJOVSKY, 1995, p. 67).

Temos assinalado na Figura 28 que a palavra *build* (construir), tomada como um verbo, possui uma estrutura argumental contendo três argumentos. O ARG_1 e ARG_2 são argumentos verdadeiros do verbo construir, visto que são realizados sintaticamente. O ARG_1 requer que sua posição seja preenchida por um **indivíduo_animado** (*animate_individual*), enquanto o ARG_2 exige que esse preenchimento se dê por um **artefato** (*artifact*). O ARG_1 marca sua posição sintática padrão como sujeito do verbo, enquanto o ARG_2 é normalmente tomado como objeto do mesmo verbo. Já o D-ARG_1 trata de um argumento *Default* ou Padrão, em que os parâmetros estão ligados à estrutura qualia que será apresentada e discutida mais adiante e não se manifesta necessariamente no âmbito sintático. Portanto o D-ARG_1 lida diretamente com o fato de que, para se construir um artefato é necessário um **material** específico (*material*).

Já a estrutura de evento adota uma visão atômica, sustentando uma representação da estrutura de subeventos associada a itens lexicais que expressam relações necessariamente entre os eventos e argumentos do verbo.

Figura 29 – Representação da estrutura de evento da palavra *build* (construir)

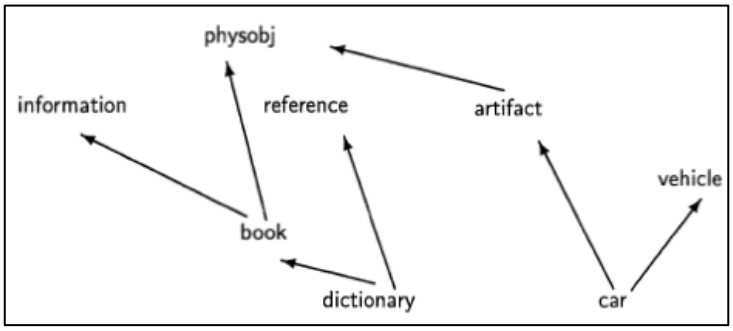


Fonte: The Generative Lexicon. (PUSTEJOVSKY, 1995, p.71).

Na Figura 29, temos a representação da estrutura de evento relacionada ao verbo *build* (construir). Percebe-se que a estrutura de evento (EVENTSTR) se desmembra em eventos menores **E_1 = processo** (*process*), **E_2 = estado** (*state*), havendo ainda uma **restrição** (<α) que demonstra o fato de poder haver uma quantidade sucessiva de eventos ordenados, os quais são subpartes do evento maior que é **construir**. O processo e E_1 seria o próprio evento que envolve a construção em si e o estado e E_2 seria a finalização esperada do evento construir, ou seja, algo que se encontra em um estado construído.

Temos também a estrutura de herança lexical que relaciona as estruturas lexicais, organizando o léxico globalmente. Essa estrutura pode ser dividida em dois tipos: a herança lexical e a herança ortogonal. A lexical é composta por uma rede estática de relações como hipônimos e hiperônimos e se assemelha a uma estrutura *lattice*. Já a herança ortogonal opera gerativamente com estrutura qualia para criar relações entre as categorias.

Figura 30 – Representação convencional da estrutura de herança lexical



Fonte: The Generative Lexicon. (PUSTEJOVSKY, 1995, p.143).

Na Figura 30, temos ilustrada a representação convencional da estrutura de herança lexical. Nota-se a relação de hipônimos e hiperônimos mencionada anteriormente ao constatarmos que o objeto **carro** (*car*) é um hipônimo de **veículo** (*vehicle*) e de **artefato** (*artifact*). **Artefato**, por sua vez, é um hipônimo de um **objeto físico** (*physobj*). Já o objeto **dicionário** (*dictionary*) se mostra como hipônimo de **livro** (*book*) e de **referência** (*reference*). **Livro**, no que lhe concerne, é um hipônimo de **objeto físico** (*physobj*) e de **informação** (*information*).

O último nível de representação semântica proposto por Pustejovsky (1995) é a estrutura qualia. Trata-se de quatro tipos de papéis, relações ou aspectos essenciais do significado das palavras. O quale constitutivo é a relação entre um objeto e suas partes constituintes. O quale formal é o que distingue um objeto dentro de um domínio maior. O quale télico se relaciona à função ou propósito do objeto. E por fim, o quale agentivo se coloca nos fatores envolvidos na origem do objeto. Essas relações foram inspiradas nos modos de explicação (*aitiae*) de Aristóteles com uma descrição rica dos significados das palavras (MORAVCSIK, 1975).

Pustejovsky e Jezek (2016) revisitam os conceitos de estrutura qualia propostos por Pustejovsky (1995) e trazem uma série de contribuições e exemplos para a compreensão do que são e como funcionam as relações qualia. Os autores propõem que

> As qualia codificam aspectos do significado de uma palavra que muitas vezes são atribuídos como conhecimento do mundo pelas teorias linguísticas

> contemporâneas, isto é, o conhecimento que temos sobre objetos no mundo devido à experiência humana. (PUSTEJOVSKY; JEZEK, 2016, p. 4)[67].

Para os autores, os significados das palavras podem ser descritos em termos de unidades menores, por uma decomposição em primitivos. Os primitivos são traços, componentes ou elementos mínimos que não podem mais ser decompostos. A Teoria do Léxico Gerativo (TLG) interpreta os significados das palavras, analisando uma ampla gama de interpretações contextuais. Os autores postulam que, para se identificar o significado de uma palavra, há a necessidade de se conhecer o sistema de representação lexical desta palavra, o que faz com que ela assuma diferentes significados conforme o contexto em que ela se encontre.

Figura 31 – Representação da Estrutura Qualia relacionada a uma entidade α

$$\begin{bmatrix} \alpha \\ \text{QUALIA} = \begin{bmatrix} \text{F} = \textbf{what } \alpha \textbf{ is} \\ \text{C} = \textbf{what } \alpha \textbf{ is made of} \\ \text{T} = \textbf{function of } \alpha \\ \text{A} = \textbf{origin of } \alpha \end{bmatrix} \end{bmatrix}$$

Fonte: Integrating Generative Lexicon and Lexical Semantic Resources. (PUSTEJOVSKY; JEZEK, 2016, p. 8).

A Figura 31 exemplifica como a estrutura qualia é representada dentro de matrizes, suas relações com uma certa entidade α e o que significam. As relações qualia são marcadas pelas iniciais de cada palavra, sendo: F (Formal, ou seja, o que é α), C (Constitutivo, isto é, de que α é feito), T (Télico, ou melhor, a função de α) e A (Agentivo, quer dizer, a origem de α).

O quale formal (*é-um*) é uma relação estabelecida entre entidades que distingue uma entidade dentro de um domínio maior. Ela traz informações taxonômicas do objeto, incluindo características como a orientação, forma, dimensões, cor, posição, tamanho etc. Cada atributo pode ser preenchido por um valor. Vejamos que em "sapato preto", "preto" é valor (preenchedor ou descritor) do atributo formal "cor". Portanto, diz-se que "preto" é formal de "sapato" na expressão "sapato preto". Essa relação pode ser demonstrada através de uma entidade denotada por uma palavra (ex.: cachorro) e a categoria à qual ela pertence (ex.: animal), distinguindo a entidade (cachorro) em um grupo maior de entidades (ex.: pedra, amor, samambaia, sapato etc.).

[67] "Qualia encode aspects of a word's meaning that are often attributed as world knowledge by contemporary linguistic theories, i.e., the knowledge we have about objects in the world due to human experience." (PUSTEJOVSKY; JEZEK, 2016, p. 4).

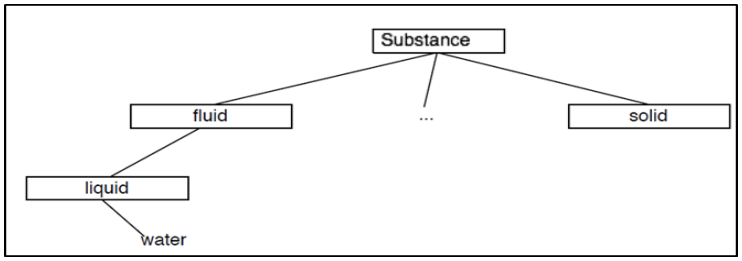
Pode-se dizer então que “animal” é formal de “cachorro”, visto que essa é uma característica inerente da entidade “cachorro” e que a diferencia de outras entidades.

Os autores resumem o papel Formal a partir dos seguintes tipos de informação:

> a. A categoria básica associada à palavra (isto é, seu tipo semântico);
> b. A posição da palavra na hierarquia dos tipos que se seguem dessa associação;
> c. As propriedades salientes que entram na definição do tipo, que são herdadas pela palavra através do papel Formal. (PUSTEJOVSKY; JEZEK, 2016, p. 16)[68].

O Quale Formal pode ser representado através de uma Hierarquia Lexical, refletindo a Estrutura de Tipos das Entidades que detalhamos anteriormente.

Figura 32 – Representação de Hierarquia Lexical refletindo a Estrutura de Tipos



Fonte: Integrating Generative Lexicon and Lexical Semantic Resources. (PUSTEJOVSKY; JEZEK, 2016, p. 14).

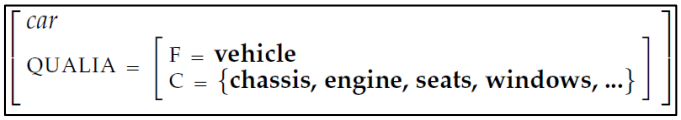
A Figura 32 traz uma esquematização do Quale Formal a partir das entidades **água** (*water*), **líquido** (*liquid*), **fluido** (*fluid*), **substância** (*substance*) e **sólido** (*solid*). Pode-se dizer que líquido, fluído e substância são valores formais da entidade água, visto que são as características que a diferenciam em domínios maiores. Com base nas propriedades que a palavra herda de seus superordenados é que se torna possível o estabelecimento de inferências das classes e suas características peculiares.

O quale constitutivo (*parte-de* ou *feito-de*) é a relação que ocorre entre um objeto e suas partes constituintes e o material envolvido. Os elementos importantes para essa relação são o material, o peso e outras partes etc. que compõem a entidade. Além das partes que constituem um objeto, é importante também se considerar o que está “dentro” dele ou de que material ele é feito.

---

[68] “a. The basic category associated with the word (i.e., its semantic type);
b. The position of the word in the hierarchy of types following from this association;
c. The salient properties which enter into the definition of the type, which are inherited by the word along the Formal role.” (PUSTEJOVSKY; JEZEK, 2016, p. 16).

Figura 33 – Representação de Qualia Formal e Constitutivo da entidade carro (*car*)



Fonte: Integrating Generative Lexicon and Lexical Semantic Resources. (PUSTEJOVSKY; JEZEK, 2016, p. 21).

A Figura 33 traz a representação de estrutura Qualia Formal e Constitutivo para a entidade carro (*car*). Observamos que veículo (*vehicle*) preenche seu Quale Formal, enquanto que chassi (*chassis*), motor (*engine*), assentos (*seats*) e janelas (*windows*) denotam as partes do carro ou preenchem o valor de Quale Constitutivo da entidade carro.

Pustejovsky e Jezek (2016) postulam que a palavra **parte**, fundamental para a construção do conceito de Constitutivo "[...] pode ser utilizada para designar qualquer porção de uma dada entidade, independentemente, por exemplo, de a porção estar ligada ao resto do objeto, 'a maçaneta de uma porta', ou desprendida do objeto, 'a tampa de uma caneta'." (PUSTEJOVSKY; JEZEK, 2016, p. 22)[69]. Os autores definem as partes como:

> a. partes disponíveis no discurso como unidades individuais;
> b. partes fazem uma contribuição funcional para a entidade;
> c. partes são cognitivamente salientes. (PUSTEJOVSKY; JEZEK, 2016, p. 22)[70].

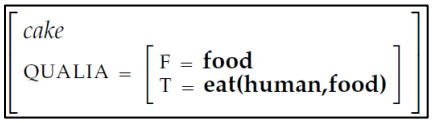
O quale télico (*usado-para* ou *funciona-como*) é a relação estabelecida entre uma entidade e seu propósito ou função inerente. Quando pensamos em objetos, o papel télico do objeto diz respeito ao que nós normalmente fazemos com aquele objeto, ou seja, qual a função daquele objeto. Quando as entidades são pessoas, temos que pensar no quale télico como as ações prototípicas que aquelas pessoas executam. Ao pensarmos em entidades relacionadas a lugares, trata-se das ações relacionadas àqueles lugares prototipicamente. Trata-se de uma propriedade persistente do objeto, ou seja, quando a atividade é realizada, o principal propósito da entidade se satisfaz. Quando lidamos com entidades que são artefatos, o papel télico se relaciona ao propósito que a entidade tem ao ser criada. (Ex.: cortar é télico de faca, isto é, a faca enquanto um artefato ou algo criado veio a existir com o propósito de cortar coisas).

---

[69] "[…] may be used to indicate any portion of a given entity, regardless of whether, for example, that portion is attached to the rest of the object, as with "the handle of a door", or undetached, as "the cap of a pen." (PUSTEJOVSKY; JEZEK, 2016, p. 22).

[70] "a. parts are available in discourse as individual units;
b. parts make a functional contribution to the entity;
c. parts are cognitively salient." (PUSTEJOVSKY; JEZEK, 2016, p. 22).

Figura 34 – Representação de Estrutura Qualia Formal e Télico da entidade bolo (*cake*)

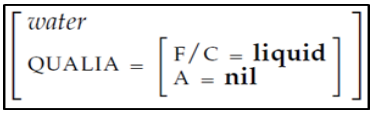


Fonte: Integrating Generative Lexicon and Lexical Semantic Resources. (PUSTEJOVSKY; JEZEK, 2016, p. 35).

A Figura 34 demonstra as relações qualia Formal e Télico para a entidade bolo (*cake*). Observamos que comida (*food*) preenche o valor formal de bolo (*cake*). Comer (*eat*) aparece como télico de bolo, ou seja, a entidade bolo tem por função o fato de alimentar enquanto comida (*food*) o ser humano (*human*) que a come.

O quale agentivo (*criado-por*) é a relação estabelecida entre uma entidade e os fatores envolvidos em sua origem, ou seja, que elementos influenciam no fato de essa entidade passar a existir no mundo. Características incluídas nessa relação são o criador, o artefato, o tipo natural e uma cadeia causal. Há entidades de origem natural que apresentam um valor zero para o papel agentivo. Assim, o valor zero (*nil*) para o papel agentivo captura a primazia de uma origem natural. Dessa forma, não se estabelece uma relação dessa entidade com outras que contribuiriam para que essa entidade se originasse. Vejamos uma exemplificação desse valor.

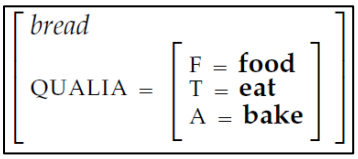
Figura 35 – Representação de Qualia Formal, Constitutivo e Agentivo (*nil*) da entidade água (*water*)



Fonte: Integrating Generative Lexicon and Lexical Semantic Resources. (PUSTEJOVSKY; JEZEK, 2016, p. 34).

A Figura 35 traz a representação das relações qualia Formal, Constitutivo e Agentivo para a entidade água (*water*). Percebe-se que os valores Formal e Constitutivo de água são preenchidos por líquido. Devido ao fato de água ser uma substância de origem natural, em que não há uma precedência em termos de origem, o valor preenchido para o papel Agentivo torna-se um valor zero (*nil*), representando um valor nulo, uma origem ou início de algo. O quale Agentivo também se relaciona ao modo de origem ou criação da entidade.

Figura 36 – Representação de Qualia Formal, Télico e Agentivo da entidade pão (*bread*)



Fonte: Integrating Generative Lexicon and Lexical Semantic Resources. (PUSTEJOVSKY; JEZEK, 2016, p. 35).

Na Figura 36, temos a representação das relações qualia para a entidade pão (*bread*). Comida (*food*) é formal de pão, comer (*eat*) é télico de pão e assar (*bake*) é Agentivo de pão, tendo em vista que o pão só foi originado a partir da ação de assar. O papel agentivo pode ser compreendido como uma pré-condição para a existência da entidade. Dessa forma, o criador da entidade também pode ser incluído como possuindo uma relação agentiva com essa entidade.

A partir de todo o detalhamento do conceito de entidade numa ontologia conceitual de representação do conhecimento proposto por Helbig (2006) e das relações qualia postuladas e discutidas por Pustejovsky (1995) e Pustejovsky e Jezek (2016), tracemos algumas comparações. No escopo deste trabalho, Entidade é sinônimo de Objeto para Helbig (2006). O tipo de relação qualia Télico_de (PUSTEJOVSKY, 1995), que se refere à função ou propósito de uma entidade, pode se traduzir como as Situações Dinâmicas (HELBIG, 2006). A relação qualia Constitutivo_de (PUSTEJOVSKY, 1995), representando parte e todo pode ser alinhada às Propriedades Relacionais (HELBIG, 2006).

Traçando uma analogia entre a TLG e a Semântica de Frames, os níveis de representação semântica propostos pela TLG se apoiam na descrição dos itens lexicais, enquanto a Semântica de Frames, por sua vez, se coloca ao nível do frame. A estrutura argumental pode ser comparada aos padrões de valência sintático-semântica (EF, FG, PT) das ULs na FrameNet. Já a estrutura de eventos se assemelha à rede de frames em si que, organizada em forma de *lattice* e ilustrada pelo *Grapher*, busca relacionar os frames partindo de frames mais genéricos e atingindo frames mais específicos. Por fim, a estrutura de herança lexical proposta se equipara às relações entre frames da FrameNet, já que, ainda que indiretamente, elas estabelecem certos tipos de relação de herança entre as palavras.

As relações qualia surgem como forma de melhorar a granularidade semântica das descrições lexicais dentro da FrameNet. Ao tratar da semântica dos nomes de entidade, devemos nos atentar para a diferenciação que ocorre na seleção argumental entre nomes e verbos. Para os verbos e sintagmas nominais que denotam eventos, a FrameNet faz a anotação dos padrões de valência, marcando o que a TLG propõe como a estrutura argumental e de evento

dos itens lexicais através na anotação lexicográfica ou de texto corrido. Já para os nomes de entidades, a semântica dos nomes (estrutura qualia e outros modelos de estruturação) proposta por Pustejovsky surgiria com a finalidade de "facilitar e possibilitar uma interpretação composicional mais rica na caracterização da semântica das línguas naturais enquanto polimórfica" (PUSTEJOVSKY, 1995, p. 141).

Acerca da tradução automática de verbos e nomes eventivos, Perón-Corrêa *et. al.* (2016) propõem que os verbos podem apresentar um desempenho satisfatório na TM porque possuem padrões de valência mais informativos. Partindo dessa análise, nos interessam muito a utilização das relações qualia para contribuir no enriquecimento semântico dos nomes de entidades, dado o fato de os substantivos não apresentarem padrões de valência tão informativos.

A modelagem semântica com estrutura qualia para as Unidades Lexicais nos permite lidar de forma eficiente com a desambiguação de nomes de entidade, melhorando a compreensão de línguas naturais e processos de tradução por máquina, diante de uma rede mais adensada de relações entre os elementos. As relações qualia trazem aos itens lexicais que representam nomes de entidades padrões de valência mais informativos que não possuíam anteriormente como os verbos.

Refletindo acerca da melhoria na granularidade semântica das descrições lexicais e representação semântica, Belcavello *et al.* (2020)[71] apresentam uma proposta de ampliação dos conceitos de Qualia propostos por Pustejovsky e Jezek (2016) ao postularem as Relações Qualia Ternárias mediadas por Frames. Tendo observado que as relações qualia propostas originalmente ainda apresentam um aspecto muito genérico, a FrameNet Brasil viu a possibilidade de "[...] usar frames nesse mesmo banco de dados como mediadores de relações qualia ternárias para abordar tanto a falta de links diretos entre ULs no modelo da FrameNet quanto a baixa especificidade das relações qualia." (BELCAVELLO *et al.*, 2020, p. 26)[72]. Duas ULs são ligadas entre si através de uma relação qualia, mas possuindo um frame de pano de fundo, o que adensa a relação qualia com informações oferecidas pelo frame que intermedia a relação. Para cada tipo de relação qualia foram escolhidos frames que poderiam intermediar a relação, estando a UL_1 associada a um EF_1 desse frame e a UL_2 relacionada a outro EF_2 do mesmo frame. Essas relações qualia ternárias ocorrem de maneira direcional e devem ser interpretadas em uma única direção, podendo novas relações e novos frames serem criados

[71] A proposta de relações qualia ternárias foi desenvolvida no âmbito desta tese, tendo sido publicada parcialmente no artigo em questão, de forma antecipada à defesa da tese.

[72] "[…] we use frames in this same database as mediators of ternary qualia relations to address both the lack of direct links between LUs in the FrameNet model and the poor specificity of qualia relations." (BELCAVELLO *et al.*, 2020, p. 26).

conforme as necessidades da pesquisa. A escolha dos frames que compõem as relações qualia ternárias ocorre através de dois critérios: (i) o frame mais genérico possível e (ii) um frame tão específico quanto necessário. A ideia é que frames genéricos consigam abarcar características da relação qualia, mas trazendo um enquadramento semântico para a relação, tornando-a mais específica. A utilização do segundo critério traz a possibilidade de escolha de um frame mais específico para uma relação com o propósito de que aspectos fundamentais daquela relação qualia não se percam caso o frame de fundo seja genérico demais. A Figura 37 traz uma tabela com cinco colunas que exemplificam todas as relações qualia possíveis de serem criadas atualmente. A primeira coluna mais à direita traz o tipo de relação qualia (Agentivo, Télico, Formal e Constitutivo). A segunda coluna traz o frame que intermedia a relação, ou seja, o frame que oferece características para que a relação qualia se torne mais específica. A terceira coluna mostra o EF_1 da relação, ou seja, o EF com o qual a UL_1 irá se associar. A quarta coluna (*info*) traz informações sobre a relação, auxiliando na leitura direcional da relação proposta. A quinta e última coluna ilustra o EF_2 da relação com o qual a UL_2 será associada. Ambos EF_1 e EF_2 precisam ser elementos nucleares do frame presente na segunda coluna, fazendo com que a relação qualia seja mais específica.

Figura 37 – Relações Qualia Ternárias na FrameNet Brasil

Type	Frame	LU1	Info	LU2
Qualia Agentive	Afetar_intencionalmente	Paciente	afetado por	Agente
Qualia Agentive	Agir_intencionalmente	Ação	causado por	Agente
Qualia Agentive	Causalidade	Efeito	causado por	Causa
Qualia Agentive	Causalidade	Efeito	causado por	Ator
Qualia Agentive	Criação_culinária	Comida_produzida	criado por	Cozinheiro
Qualia Agentive	Criar_intencionalmente	Entidade_criada	criado por	Criador
Qualia Agentive	Inovar	Nova_ideia	criado por	Pensador
Qualia Agentive	Resolver_problema	Problema	resolvido por	Agente
Qualia Constitutive	Influência_objetiva	Entidade_influenciadora	afeta	Entidade_dependente
Qualia Constitutive	Causalidade	Ator	causa	Afetado
Qualia Constitutive	Conter	Recipiente	contém	Conteúdos
Qualia Constitutive	Ingredientes	Produto	feito com	Material
Qualia Constitutive	Inclusão	Total	inclui	Parte
Qualia Constitutive	Empregar	Empregador	local de trabalho de	Empregado
Qualia Constitutive	Parentesco	Ego	parente de	Alter
Qualia Constitutive	Criar	Criador	produz	Entidade_criada
Qualia Constitutive	Agir_intencionalmente	Ação	realizado por	Agente
Qualia Constitutive	Relação	Entidade_1	relaciona-se com	Entidade_2
Qualia Constitutive	Pessoas_por_religião	Pessoa	seguidor de	Religião
Qualia Constitutive	Atributos	Entidade	tem como atributo	Atributo
Qualia Constitutive	Associação	Grupo	tem como membro	Membro
Qualia Constitutive	Parte_interior_exterior	Todo	tem como parte	Parte
Qualia Constitutive	Parte_todo	Todo	tem como parte	Parte
Qualia Constitutive	Parte_elemento	Substância	tem como parte	Elemento
Qualia Constitutive	Subpartes_de_prédios	Todo	tem como parte	Parte
Qualia Constitutive	Residência	Local	tem como residente	Residente
Qualia Constitutive	Pessoas_por_origem	Pessoa	tem origem em	Origem
Qualia Constitutive	Usar_recurso	Agente	usa	Recurso
Qualia Constitutive	Infraestrutura	Infraestrutura	utilizado por	Usuário
Qualia Formal	Exemplar	Exemplar	exemplo de	Tipo
Qualia Formal	Tipo	Subtipo	tipo de	Categoria
Qualia Telic	Agir_intencionalmente	Ação	atividade de	Agente
Qualia Telic	Criar_intencionalmente	Entidade_criada	criado por	Criador
Qualia Telic	Finalidade_do_utensílio	Finalidade	finalidade de	Utensílio
Qualia Telic	Capacidade_ação	Evento	habilidade de	Entidade
Qualia Telic	Costume	Comportamento	hábito de	Protagonista
Qualia Telic	Finalidade	Alvo	objetivo de	Agente
Qualia Telic	Infraestrutura	Atividade	realizado em	Infraestrutura
Qualia Telic	Usar	Agente	utilizado para	Propósito
Qualia Telic	Usar_recurso	Recurso	utilizado por	Agente
Qualia Telic	Vício	Vício	vício de	Viciado

Fonte: Webtool da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

A Figura 37 apresenta os tipos de relações qualia ternárias possíveis de serem estabelecidas entre ULs na FrameNet Brasil atualmente. As quatro relações mais genéricas (Agentivo, Constitutivo, Formal e Télico) passam então a ser incorporadas na base de dados da FrameNet Brasil como quarenta e uma relações qualia mais específicas ternárias e mediadas

por frames. Trata-se de oito tipos de relação do tipo Agentivo, vinte e uma relações do tipo Constitutivo, duas relações do tipo Formal e dez relações do tipo Télico.

Analisemos agora um exemplo de modelagem específica de relações qualia ternárias para o domínio dos Esportes.

Figura 38 – Exemplos de Relações Qualia ternárias modeladas na FrameNet Brasil

Type	LU1	Relation	LU2
agentive	dois toques.n	causado por	jogador de vôlei.n
constitutive	quadra de vôlei.n	tem como parte	linha de saque.n
formal	jogador de vôlei.n	tipo de	jogador.n
telic	cortada.n	atividade de	jogador de vôlei.n

Fonte: Webtool da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

A Figura 38 exibe alguns exemplos de relações qualia ternárias modeladas no domínio dos Esportes na FrameNet Brasil. O primeiro exemplo que temos é de uma relação agentiva entre a UL_1 **dois toques.n** e a UL_2 **jogador de vôlei.n**. A relação agentiva nesse exemplo é mediada pelo frame `Agir_intencionalmente`, estando a UL_2 associada ao EF_2 Agente e a UL_1 relacionada ao EF_1 Ação. Portanto, **jogador de vôlei.n** é agentivo de **dois toques.n**, sendo que esta infração só passa a existir no mundo se o **jogador de vôlei.n** a realizar. No segundo exemplo, temos que a UL_1 **quadra de vôlei.n** tem constitutivamente como parte a UL_2 **linha de saque.n**. Assim, através do frame `Subpartes_de_prédios`, a UL_2 se associa ao EF_2 Parte e a UL_1 se relaciona ao EF_1 Todo. Portanto, a **linha de saque.n** é constitutiva de **quadra de vôlei.n**. No terceiro exemplo, a UL_1 **jogador de vôlei.n** tem como formal a UL_2 **jogador.n**. A partir de uma relação Formal mediada pelo frame `Tipo`, a UL_1 se relaciona ao EF_1 Subtipo, enquanto a UL_2 se associa ao EF_2 Categoria. Portanto, **jogador.n** é formal de **jogador de vôlei.n**, sendo este um subtipo daquele. No quarto e último exemplo, temos a relação télica entre a UL_1 **cortada.n** e a UL_2 **jogador de vôlei.n**. Essa relação télica é mediada pelo frame `Agir_intencionalmente`, sendo que a UL_1 se associa ao EF_1 Ação e a UL_2 se relaciona ao EF_2 Agente. Assim, a **cortada.n** é télico de **jogador de vôlei.n**, pois é uma atividade prototípica desempenhada por esse tipo de jogador.

Passemos ao Capítulo 4 em que apresentaremos o aplicativo m.knob (*Multilingual Knowledge Base*), suas funções (Sistema de Recomendação através de um *Chatbot*, Diciopédia e Intérprete Pessoal), além de compilarmos os dados utilizados na estrutura da aplicação.

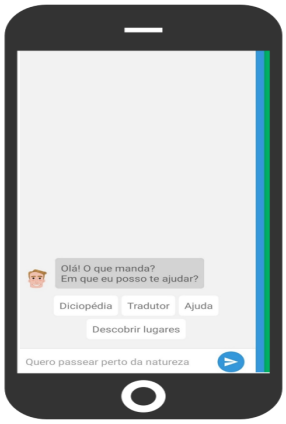
4. MULTILINGUAL KNOWLEDGE BASE (M.KNOB)

O Capítulo 4 apresenta o aplicativo m.knob (*Multilingual Knowledge Base*), ilustra a modelagem de frames, relações EF-Frame, de equivalência de tradução e qualia desenvolvida na base de dados da FrameNet Brasil, e discorre acerca de outras bases de conhecimento, ou ontologias, utilizadas pela FrameNet Brasil no desenvolvimento de suas aplicações.

4.1 VISÃO GERAL DO APLICATIVO

O m.knob (*Multilingual Knowledge Base*)[73] é em uma aplicação desenvolvida pela FrameNet Brasil na Universidade Federal de Juiz de Fora. Sua principal função é servir ao usuário como um guia turístico multilíngue de bolso que abrange os domínios do Turismo e dos Esportes. Através de um Sistema de Recomendação via *chatbot*, uma Diciopédia e um Intérprete Pessoal, ele relaciona o processamento de língua natural com ontologias e dados ligados, com o propósito de inovar a experiência turística e o uso de tradutores por máquina ou intérpretes pessoais.

Figura 39 – Interface do Aplicativo m.knob e sua tela de apresentação



Fonte: Aplicativo m.knob. (http://mknob.com).

Na Figura 39, observamos a tela inicial do aplicativo m.knob. Há um personagem em forma de um avatar (GregBot) que interage com o usuário diretamente através de um *chatbot*. Um *chatbot* pode ser compreendido como um sistema computacional que é programado para desempenhar determinadas tarefas e simular uma conversa humana através de um *chat*

[73] Acessível em: http://mknob.com.

(MAULDIN, 1994). Na tela inicial esboçada, Greg pergunta como pode ajudar o usuário, sugerindo três opções de possíveis ações que ele pode executar: (i) Descobrir lugares, (ii) Diciopédia e (iii) Tradutor, além de um botão de ajuda caso o usuário necessite de ajuda ou de interação com a equipe de desenvolvimento. O usuário também pode digitar algo na barra horizontal inferior e, submetendo alguma sentença, ele obtém alguma resposta do Greg, que representa o sistema do m.knob como um todo.

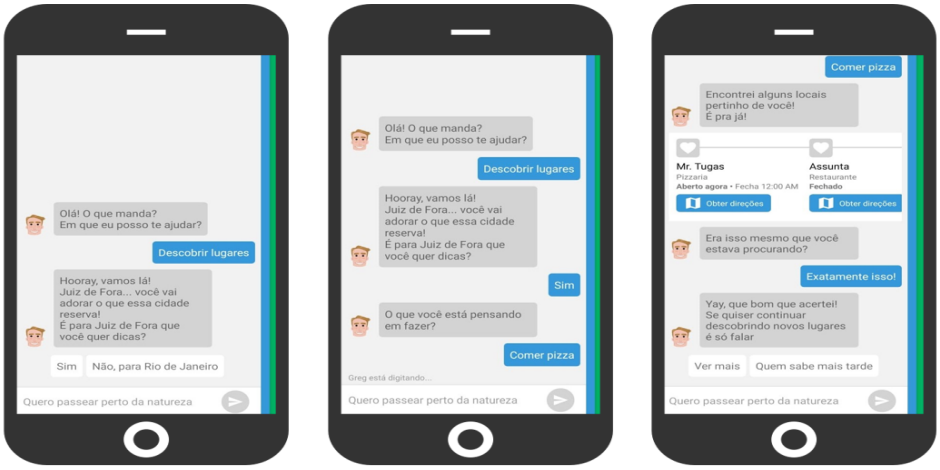
4.1.1 Função Sistema de Recomendação

A função do Sistema de Recomendação (PAIVA, 2019) do m.knob nada mais é do que um *chatbot* que interage com o usuário através do personagem GregBot,

> [...] pergunta ao usuário o que ele quer fazer e, dada uma resposta em língua natural, fornece recomendações de locais turísticos que aderem à resposta do usuário. A ideia é estimular o usuário a detalhar a sua necessidade, de modo a obter mais subsídios para a recomendação. (PAIVA, 2019, p. 124).

O sistema de recomendação se apoia em duas bases de dados, sendo elas: (i) a FrameNet Brasil, que oferece toda uma modelagem de Frames, Unidades Lexicais e relações Qualia modeladas nos frames de Turismo e Esportes; e (ii) uma base de dados dentro do aplicativo que é composta por dados dos locais turísticos das duas cidades cobertas pelo aplicativo, Juiz de Fora e Rio de Janeiro, extraídos a partir da API do *Google Places*. Vejamos agora um exemplo de uso do sistema de recomendação a partir de uma busca e interação direta com o Gregbot.

Figura 40 – Sistema de Recomendação do m.knob



(a) **(b)** **(c)**

Fonte: Aplicativo m.knob. (http://mknob.com).

A Figura 40 traz como exemplo de uso do sistema de recomendação três telas do m.knob, sendo uma a sucessão da outra na interação com o Gregbot. O usuário inicia sua interação em 40a, clicando na opção **Descobrir lugares** do m.knob. O sistema pergunta se as recomendações turísticas que o usuário quer são para as cidades de Juiz de Fora ou do Rio de Janeiro. As cidades foram escolhidas inicialmente como duas opções possíveis dado o fato de o Rio de Janeiro ser uma cidade turística mundialmente famosa e a cidade de Juiz de Fora - MG ser onde se encontra o laboratório FrameNet Brasil no qual a aplicação vem sendo desenvolvida. Em 40b, vemos que o usuário escolheu Juiz de Fora. Em seguida, o Gregbot pergunta o que o usuário deseja fazer, obtendo como resposta "**comer pizza**". O sistema utiliza a modelagem linguístico-computacional realizada das ULs e relações qualia na FrameNet Brasil. A partir de "**comer.v**" e "**pizza.n**" e suas relações modeladas, ele faz uma busca nos possíveis lugares dada a extração da API do *Google Places*. Em 40c, ele oferece como possíveis locais **Mr. Tugas** e **Assunta** que são estabelecimentos em Juiz de Fora que vendem pizza. Conclui-se assim que ele sugeriu locais adequados para as atividades turísticas que o usuário estava buscando. Abaixo do nome do estabelecimento, ele fornece informações como o tipo de estabelecimento (pizzaria, restaurante), o horário de funcionamento e um botão de obter direções para caso o usuário desejar olhar o endereço. Essas informações são extraídas do *Google Places*. Por último, após a recomendação ser feita, o sistema pergunta ao usuário se era aquilo mesmo que ele estava procurando e se oferece para ajudar em algo mais.

Paiva (2019) propõe um *chatbot* que oferece informações mais específicas ao usuário a partir de uma modelagem semântica (frames e qualia). No âmbito de seu trabalho, houve a ampliação de 37 para 134 frames considerados no domínio do Turismo, além da modelagem de 3000 relações qualia no mesmo domínio. O material textual que serviu de base para o Sistema de Recomendação foi composto pelos dados de input do usuário e um corpus de referência com comentários extraídos do Google Places. Os frames elencados do corpus foram divididos em frames primários e secundários, sendo os primários diretamente relacionados ao Turismo e os secundários aqueles que não eram diretamente ligados ou turismo ou apenas parcialmente associados. Quanto mais comentários um dado lugar possuía, maiores eram as chances de o Sistema de Recomendação sugerir esse local turístico dada a relevância da extração de informações textuais e a análise semântica do corpus.

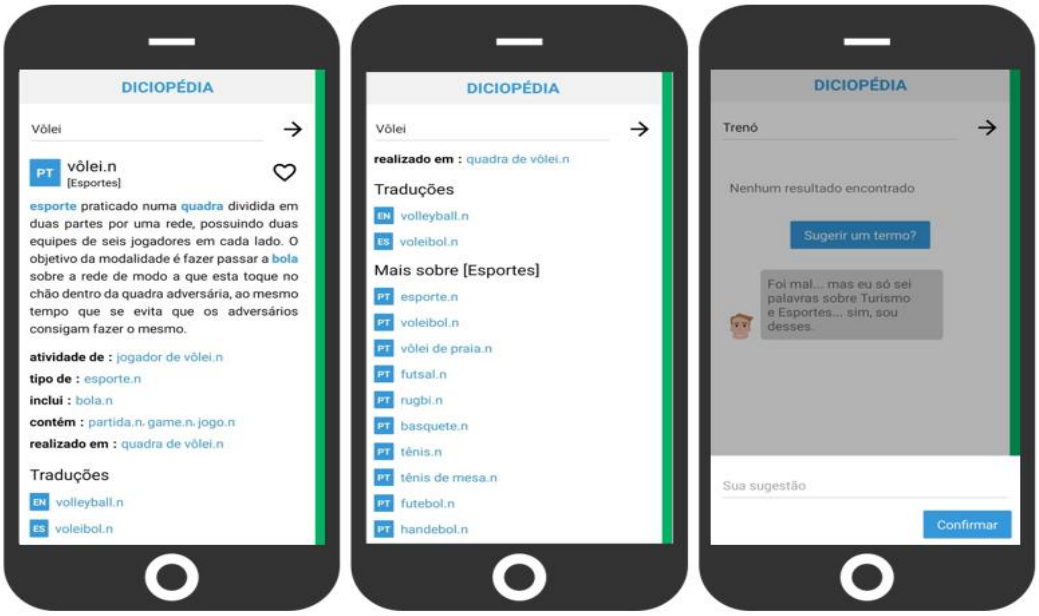
As relações qualia ternárias contribuíram, a partir das ULs que estavam presentes nos comentários do Google Places e no input do usuário, para que o sistema de recomendação apresentasse um melhor desempenho no que diz respeito às sugestões de lugares turísticos que

continham nomes de entidade. Vejamos agora como funciona a Diciopédia, outra função do m.knob.

4.1.2 Função Diciopédia

A Diciopédia é um repositório multilíngue de palavras e conceitos relacionados aos domínios do Turismo e dos Esportes dentro do m.knob (PERON-CORRÊA, 2019). Através dela, o usuário pode acessar termos do vocabulário do turismo e dos esportes em português brasileiro, inglês e espanhol. Durante a interação, pode haver a sugestão por parte de quem utiliza o aplicativo de novas palavras, além de ver relações com outras palavras (via relações qualia modeladas), traduções do termo pesquisado em uma das três línguas do app e palavras relacionadas. As traduções contidas e apresentadas no interior da Diciopédia são geradas automaticamente a partir da modelagem feita na base de dados da FrameNet Brasil. A Diciopédia pode ser acessada no m.knob de duas formas, como observado na Figura 39, sendo uma delas clicando na opção **Diciopédia** apresentada pelo Gregbot e a outra deslizando a tela para a esquerda na aba de cor azul. Analisemos em detalhes um verbete da Diciopédia e a tela de sugestão de novos termos.

Figura 41 – Diciopédia no m.knob



(a) **(b)** **(c)**

Fonte: Aplicativo m.knob. (http://mknob.com).

Na Figura 41, temos a Diciopédia ilustrada através da pesquisa do verbete **vôlei.n** (39a). Temos a palavra vôlei seguida de **.n** que é sua classe de palavra (*noun*, nome em inglês). Abaixo dela, temos o frame no qual a UL está modelada [`Esportes`]. Continuando, vemos a definição do verbete, a partir da qual há links que o usuário pode acessar para ir a outras palavras da Diciopédia. Após a definição, temos as palavras com as quais a palavra vôlei se relaciona. Essas relações são as relações qualia. Por exemplo, em 41a, vemos que **vôlei.n** é uma atividade do **jogador de vôlei.n**. Abaixo dessa relação, vemos relações com outras ULs da base. Todas as relações qualia entre ULs do domínio dos Esportes foram modeladas no âmbito desta tese. Após essas relações com **vôlei.n**, observamos duas traduções possíveis, sendo ***volleyball*.n** em inglês [en] e ***voleibol*.n** em espanhol [es]. A 41b mostra a continuação da consulta pelo verbete vôlei, e outras palavras dentro do mesmo frame Esportes no qual a UL **vôlei.n** se encontra. Por último, a 41c traz o exemplo de um usuário que digitou trenó e o sistema não gerou resultados. Surge o Gregbot novamente dizendo que possui apenas palavras de Turismo e Esportes e um botão de **Sugerir um termo**, em que o usuário pode digitar o termo e confirmar o envio.

Peron-Corrêa (2019, p. 179) propõe em sua tese "[...] um método replicável de avaliação da usabilidade de um produto lexicográfico". A autora avaliou o desenho da interface proposta na Diciopédia, verificando e quantificando aspectos relevantes e úteis ao usuário durante a interação com um dicionário eletrônico temático multilíngue. Um estudo teórico e metodológico foi realizado nas áreas de lexicografia, terminologia e aplicativos que utilizam essas teorias com o propósito de trazer para a interface elementos que podem contribuir para que o usuário tenha uma experiência mais rica em termos de consulta a dicionários e enciclopédias.

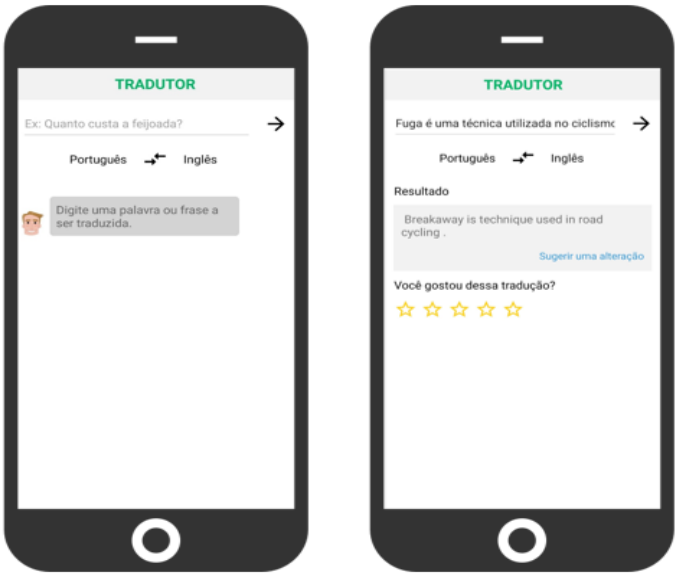
Passemos então à última função do m.knob que seria a de Intérprete Pessoal, função esta que é o foco desta tese, dado o fato de estarmos criando um tradutor de domínio específico semanticamente melhorado com frames e relações qualia.

4.1.3 Função Intérprete Pessoal

Apresentamos a última função do m.knob, o Intérprete Pessoal, alvo de pesquisa desta tese. Trata-se de um tradutor por máquina, semanticamente melhorado com frames e relações qualia, que oferece ao usuário traduções de sentenças nos domínios do aplicativo, no caso, o Turismo e os Esportes. O objetivo principal desta função do m.knob é tornar conveniente para o usuário, turista ou interessado por turismo ou esportes, enquanto recebe recomendações e descobre definições e relações entre as palavras, a utilização de um tradutor semanticamente

melhorado para atingir seu objetivo na experiência turística e/ou esportiva. A função de Intérprete Pessoal pode ser acessada como vimos na Figura 39 clicando em **Tradutor**, uma das opções do Gregbot, ou deslizando a tela duas vezes para a esquerda, na aba verde. O detalhamento do Modelo de Tradução aplicado na função Intérprete Pessoal e os testes do sistema de tradução constarão no Capítulo 6 desta tese. Vejamos a tela da função de Intérprete Pessoal.

Figura 42 – Função de Intérprete Pessoal do m.knob



(a) **(b)**

Fonte: Aplicativo m.knob. (http://mknob.com).

A Figura 42 traz as telas que representam a funcionalidade de Intérprete Pessoal ou Tradutor do m.knob. Na Figura 42a, há um espaço para que o usuário digite a sentença que ele deseja ver traduzida, o par de línguas que ele quer que o sistema traduza, podendo inverter a ordem e uma sentença de sugestão inicial a ser traduzida (Quanto custa a feijoada?). Em 40b, há uma sentença específica do domínio dos Esportes que seria "Fuga é uma técnica utilizada no ciclismo de estrada". Sistemas atuais de tradução por máquina traduzem essa sentença como "*Escape is a technique used in road cycling*". Entretanto, "escape" não seria o melhor equivalente em inglês para o nome da técnica do ciclismo "fuga" em português. Portanto, o nosso sistema oferece uma tradução semanticamente melhor "*Breakaway is a technique used in road cycling*" em que "*breakaway*" seria o equivalente mais adequado para "fuga" no domínio dos esportes. Essa tradução pode ser vista ainda na Figura 42b. A tradução específica e melhorada para os domínios do Turismo e dos Esportes se coloca como um diferencial do

m.knob para o turista ou usuário, o que traz experiências no turismo ou nos esportes mais ricas que o usuário não encontra ao utilizar sistemas de tradução tradicionais. Após visualizar a tradução oferecida, o usuário pode avaliar a tradução em até cinco estrelas e também sugerir uma nova tradução clicando em Sugerir uma alteração.

Vejamos na seção seguinte a quantificação da base de dados e como se deu o detalhamento linguístico (frames, ULs e relações qualia) utilizado na estrutura de dados do m.knob.

4.2 ESTRUTURA DA BASE DE CONHECIMENTO

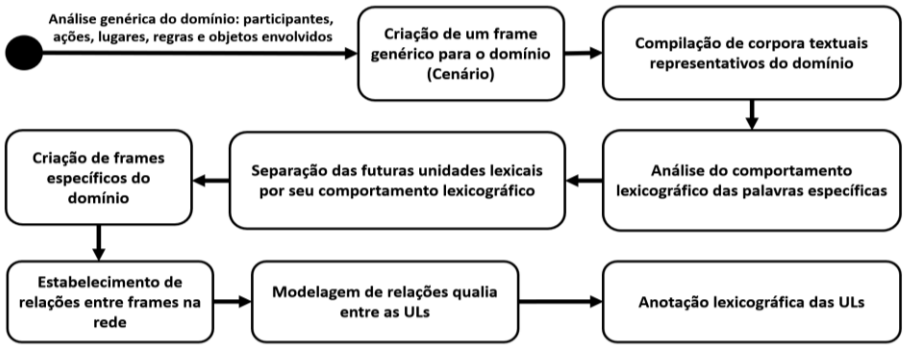
Nesta seção, abordaremos os dados que foram modelados na base da FrameNet Brasil e foram utilizados pelo m.knob. Toda a modelagem descrita aqui foi realizada no âmbito desta tese. Passaremos pela modelagem dos frames do domínio dos Esportes, as ULs criadas em português do Brasil, inglês e espanhol, além das relações qualia e buscas de palavras em outras bases de dados como a BabelNet.

Os dados relacionados ao Turismo foram modelados em uma primeira aplicação desenvolvida pela FrameNet Brasil, o dicionário Copa 2014. Torrent *et al*. (2014b) criaram e lançaram um dicionário eletrônico trilíngue (Português do Brasil, Inglês e Espanhol) para a Copa do Mundo de 2014. O dicionário era baseado em frames e cobria os domínios do Futebol, Turismo e Copa do Mundo. Toda a modelagem realizada para o domínio do Turismo (GAMONAL, 2013; GOMES, 2014; SOUZA, 2014) foi desenvolvida para o dicionário Copa e aproveitada com algumas modificações para sua inserção no m.knob. A base de dados, no domínio do Turismo, incluindo frames genéricos com relações aos frames do mesmo domínio, contava com aproximadamente 45 frames e 126 unidades lexicais. Como já se apontou, o trabalho de Paiva (2019) ampliou esse quantitativo para 62 frames e 1483 ULs.

4.2.1 Modelagem do Domínio dos Esportes

A modelagem linguístico-computacional de domínios específicos com frames, relações qualia, unidades lexicais e suas definições pode ser realizada a partir dos passos elencados no fluxograma ilustrado na Figura 43.

Figura 43 – Passos para modelagem de domínios específicos



Fonte: Elaborado pelo autor (2020).

A Figura 43 ilustra os passos para a modelagem de uma rede de frames e relações qualia para um dado domínio específico. Primeiramente, parte-se de uma análise genérica do domínio. Há uma reflexão acerca dos participantes, ações, lugares, regras, objetos, entre outros elementos que compõem o domínio. Em um segundo momento, cria-se um frame genérico do domínio, ou seja, um cenário que irá organizar os frames pertencentes ao domínio específico. Na terceira etapa, compilam-se corpora com textos pertencentes a gêneros representativos do domínio. Então, analisa-se o comportamento lexicográfico de palavras específicas. Dessa forma, começa-se a projetar as futuras unidades lexicais que evocam frames nesse domínio. Realiza-se uma separação dessas prováveis ULs conforme seu comportamento sintático-semântico instanciado pelas sentenças dos corpora. A intuição do linguista é fundamental para a separação das ULs e a criação dos frames de domínio específico. A partir do agrupamento dessas ULs com comportamento lexicográfico semelhante ou próximo, criam-se frames, com suas definições, EFs nucleares e não-nucleares. Parte-se para a ligação entre esses frames e com outros frames da rede da framenet através de relações frame-a-frame. Modelam-se as ULs nos frames que elas evocam a partir dos contextos extraídos dos corpora, com uma definição para cada UL. Após a criação das ULs, as relações qualia entre elas e com outras ULs da base de dados são estabelecidas. Por fim, há a anotação lexicográfica de sentenças que contenham essas ULs de modo a representar o comportamento sintático-semântico das ULs em contexto.

A modelagem do domínio dos esportes começa pelo frame mais genérico e de organização do domínio `Cenário_do_Esporte`. A rede de frames que o compõe foi criada especificamente durante a elaboração deste trabalho. Englobando os frames genéricos ligados aos frames desse domínio, a base de dados desse domínio conta com aproximadamente 36 frames em português do Brasil, inglês e espanhol, sendo 7 frames genéricos incorporados da base da FrameNet Brasil e 29 frames específicos dos esportes criados nesta pesquisa. Dentro

desses frames foram modeladas e ligadas via relações de equivalência 1.651 Unidades Lexicais em português, 2.051 ULs em inglês e 1.059 ULs em espanhol.

O processo de criação e modelagem dos frames surge a partir da compilação de corpora de textos relacionados aos esportes para o português durante o mestrado. Os textos foram extraídos do site Portal Brasil 2016[74], criado para os Jogos Olímpicos Rio 2016, bem como de manuais de esportes e sites de associações brasileiras e de notícias sobre cada esporte. Durante o doutorado, há um adensamento da base de dados através da criação de mais ULs dos esportes para o português, e a modelagem de ULs para o inglês e espanhol a partir de sites oficiais, de associações e federações esportivas e manuais específicos dos esportes. Os manuais utilizados para o português foram o "Almanaque Olímpico 2016" (FREITAS; BARRETO, 2016) e "O Livro dos Esportes: os Esportes, as Regras, as Táticas e as Técnicas" (RODRIGUES; NUNO; SALERNO, 2012). Para a modelagem das ULs dos esportes em inglês, foram utilizados os manuais "The Visual Dictionary of Sports and Games" (CORBEIL; ARCHAMBAULT, 2009), "Dictionary of Sports and Games Terminology" (ROOM, 2010) e "The Sports Book: the Games, the Rules, the Tactics and the Techniques" (BRIDLE *et al.*, 2011). Já para o espanhol, os manuais esportivos utilizados foram "El Libro de los 1001 porqués de los Deportes" (SALVA, 2011) e "Enciclopedia Visual de los Deportes" (FORTIN, 2008). Os manuais, sites de associações esportivas e sites com notícias contribuíram para uma análise linguística do uso em textos reais de possíveis unidades lexicais do domínio dos Esportes. Os corpora dos esportes servem para atestar lexicograficamente as ocorrências de possíveis novas ULs de domínio específico. Não se faz necessário um balanceamento estatístico refinado nesta etapa. A Tabela 2 traz a compilação de corpora dos esportes utilizada na modelagem em número de palavras para cada gênero textual relacionado aos esportes.

Tabela 2 - Constituição de Corpora dos Esportes (Em número de palavras)

Gêneros Textuais sobre Esportes	**Português**	**Inglês**	**Espanhol**
Notícias	596.034	173.612	113.700
Descrições	155.940	35.565	60.162
Sites governamentais	19.022	18.259	18.982
Manuais	25.347	20.314	20.386
Total	**796.343**	**247.750**	**213.230**

Fonte: Elaborado pelo autor (2020).

[74] http://www.brasil2016.gov.br/pt-br

A ferramenta de compilação de corpora utilizada na pesquisa de termos é o SketchEngine, previamente mencionado. Após a compilação dos dados, da experiência com a FrameNet e Semântica de Frames e do estudo detalhado do domínio dos esportes a partir da leitura de manuais, sites de associações esportivas e de notícias sobre esportes, os dados foram linguisticamente analisados e agrupados inicialmente em uma planilha. O agrupamento dos termos foi efetuado com uma investigação linguística minuciosa dos dados. Esse estudo foi fundamentado em observações realizadas no WordSketch dentro do SketchEngine. O comportamento sintático-semântico das possíveis ULs dos esportes foi levado em consideração diante da ocorrência das palavras no corpus e como ocorrem nas sentenças.

Figura 44 – Pesquisa da possível UL saltar em corpora no SketchEngine

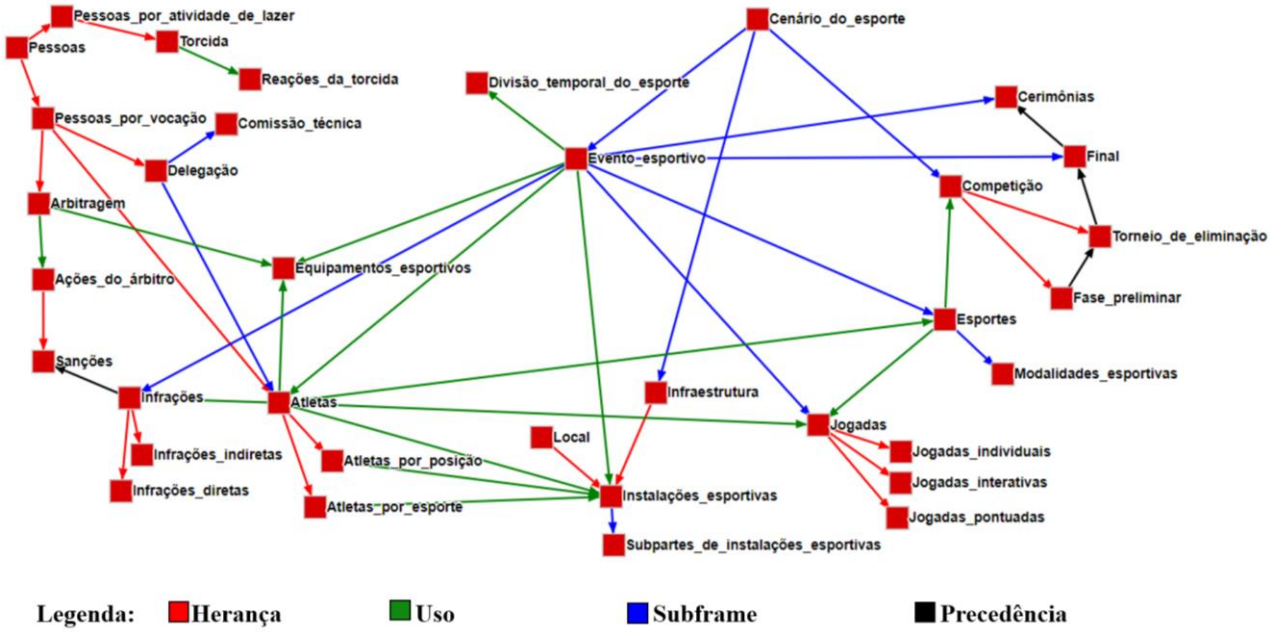
Fonte: Ferramenta de compilação de corpora SketchEngine. (https://the.sketchengine.co.uk).

A Figura 44 traz as ocorrências do termo e possível UL **saltar.v** no corpus. Ao se analisar o comportamento lexicográfico dessa UL, chega-se à conclusão de que ela se enquadraria em um frame chamado `Jogadas_individuais`, visto que não pressupõe necessariamente a interação com outro atleta e não implica pontuação direta a partir da ação em si. O processo de criação e diferenciação entre os frames para jogadas (jogadas individuais,

interativas e pontuadas) se deu com uma análise linguística minuciosa realizada no âmbito desta pesquisa. Posteriormente a isso, os dados foram modelados na criação de frames, EFs e alocação das ULs conforme o frame que evocam, além do estabelecimento das relações que ocorrem entre esses frames. Partindo da modelagem das ULs nos frames, realiza-se a anotação de sentenças que atestam sua ocorrência em corpus. Essas sentenças anotadas sintático-semanticamente foram extraídas do corpus compilado previamente no *SketchEngine*.

Vejamos o gráfico gerado na ferramenta *Grapher* na *Webtool* da FrameNet Brasil com todos os frames criados para o `Cenário_do_esporte` na Figura 45.

Figura 45 – Frames que compõe o `Cenário_do_esporte`



Fonte: Ferramenta *Grapher* na *Webtool* da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

Inicialmente chegou-se no seguinte agrupamento de categorias que se tornariam os frames dos esportes: pessoas, jogadas, regras, instalações, equipamentos, etapas e pontuação. Tal agrupamento é oriundo da sistematização do conhecimento levantado pela equipe de bolsistas de iniciação científica do projeto acerca de cada uma das modalidades e submodalidades esportivas incluídas nos jogos olímpicos de verão. Utilizou-se também a experiência acumulada na modelagem do evento da Copa do Mundo, realizada no projeto anterior.

A observação dos dados é muito importante nessa fase da pesquisa, pois é devido às semelhanças e diferenças entre as ULs que surgem as formas de agrupamento, mantendo-se

algumas categorias anteriormente pensadas e excluindo-se outras. Parte-se de uma abordagem *bottom-up* em que as ULs são elencadas e em seguida agrupadas dentro dos frames criados. A partir daí, se elaboram as definições para esses frames, seus Elementos de Frame e as relações entre eles. Foram 7 frames incorporados ao domínio dos Esportes. Os frames `Competição`, `Local`, `Infraestrutura`, `Pessoas`, `Pessoas_por_vocação` e `Vestuário` já existiam na FrameNet Berkeley, tendo sido traduzidos e incorporados na rede semântica do m.knob, visto que se relacionam com os frames criados para o domínio de esportes. Já o frame `Pessoas_por_atividade_de_lazer` foi criado pela FrameNet Brasil e também incorporado, não exclusivamente, ao domínio dos Esportes. Outros 29 frames exclusivos para os esportes foram criados e podem ser acessados na Webtool da FrameNet Brasil[75]. Detalharemos então os frames e as relações estabelecidas entre eles.

Refletindo acerca das pessoas envolvidas nos esportes, temos o frame pai `Pessoas` que possui uma relação de herança (*Inheritance*) simbolizada pelas setas na cor vermelha com os frames filhos `Pessoas_por_atividade_de_lazer` (que inclui figuras como turista, visitante, viajante, jogador) e `Pessoas_por_vocação` (que envolve profissões das mais diversas). Herdando de `Pessoas_por_atividade_de_lazer` temos `Torcida` (quem assiste ao evento). O frame de `Torcida` possui uma relação de Uso (*Using*), marcada pela seta na cor verde, com o frame `Reações_da_torcida`. Herdando de `Pessoas_por_vocação`, ressaltemos `Delegação` (quem compete ou auxilia os competidores do evento), `Arbitragem` (quem organiza e orienta o evento em termos de regras) e `Atletas` (competidores do evento). O frame `Delegação` se desmembra pela relação de subframe (*Subframe*), ilustrada pela seta na cor azul, em outros dois frames, `Comissão_técnica` (técnicos, médicos e profissionais que acompanham os atletas) e `Atletas` (termos genéricos utilizados para designar os indivíduos que praticam um esporte profissionalmente). Ao se analisarem as ULs relacionadas ao frame `Atletas` e seu comportamento lexicográfico, evidenciou-se a necessidade de um detalhamento e uma especificação do frame que, por herança, se desmembra em outros dois: `Atletas_por_esporte` (todo atleta que incorpora o nome do esporte na constituição de seu nome, tenista, por exemplo) e `Atletas_por_posição` (atletas descritos com base na posição em que ocupam, lateral, por exemplo).

[75] http://webtool.framenetbr.ufjf.br/

O frame Atletas possui uma relação de Uso com o frame Jogadas. O frame Jogadas (mais genérico designando as ações desempenhadas pelos atletas durante a prática de um esporte) se decompõe por herança e pelo comportamento lexicográfico das ULs em três frames filhos distintos: Jogadas_individuais (jogadas realizadas pelo Atleta com foco na jogada em si ou na técnica utilizada), Jogadas_interativas (jogadas realizadas por um atleta em sua interação com outros atletas) e Jogadas_pontuadas (jogadas realizadas pelo atleta ou equipe em que a jogada garante pontos para si).

Tomando o frame Arbitragem, temos a relação de uso com outros dois frames: Equipamentos_esportivos (objetos utilizados em uma competição esportiva, incluindo os aparelhos e objetos utilizados pelos atletas e pelos árbitros) e Ações_do_árbitro (ações do árbitro durante o jogo, as quais consistem em fazer a partida ocorrer conforme as regras oficiais). Há uma relação de herança entre Ações_do_árbitro e seu frame filho Sanções (punição ou pena aplicada a um atleta ou equipe ou em favor do adversário, devido a uma infração de uma regra do esporte). Na mesma linha das regras, temos o frame de Infrações (mais genérico em que uma irregularidade marcada pelo árbitro impõe uma sanção sobre um atleta ou equipe, sendo uma vantagem ao adversário), que possui uma relação de uso a partir de Atletas. Através do comportamento lexicográfico deste frame, viu-se a necessidade de um refinamento dele em outros dois frames filhos por relação de herança: Infrações_diretas (infração de um infrator contra um adversário, considerada imprudente, temerária ou com o uso de força excessiva) e Infrações_indiretas (infração em que um infrator viola uma regra do esporte que não envolva força excessiva contra o oponente). O frame de Infrações possui uma relação de Precedência (Precedes), representada pela seta na cor preta, com o frame de Sanções, visto que as ULs que representam as infrações precedem temporalmente as ULs das sanções aplicadas.

Refletindo acerca dos locais envolvidos com os esportes, criou-se o frame Instalações_esportivas (prédio, estrutura ou lugar em que uma competição de um determinado esporte ocorre) que herda do frame Local e tem uma relação de subframe com seu frame filho Subpartes_de_instalações_esportivas (objetos pertencentes à instalação esportiva ou partes das instalações esportivas).

O frame Cenário_do_esporte (cena em que um atleta pratica algum esporte, visando ao bem-estar ou a vencer uma competição regida por regras e fazendo uso de determinados equipamentos esportivos e instalações esportivas) tem como subframe o Evento_esportivo (eventos relacionados à prática de um esporte). O frame

Evento_esportivo possui uma relação de subframe com os frames Esportes (formas de praticar atividade física que, através de participação ocasional ou organizada, visa equilibrar a saúde ou melhorar a aptidão física e/ou mental, proporcionar entretenimento e/ou gerar competição entre os participantes), Jogadas, Infrações, Final e Cerimônias. Possui também uma relação de uso com os frames Atletas e Equipamentos_esportivos. O frame Esportes possui da mesma forma uma relação de subframe com o frame Modalidades_esportivas (subcategorias esportivas ou tipos de modalidade do esporte).

Ponderando sobre as expressões de tempo que se relacionam às etapas de uma partida esportiva, temos que o frame Evento_esportivo (partidas ou palavras análogas em cada esporte) possui uma relação de uso com o frame Divisão_temporal_do_esporte (tempos ou etapas de uma partida esportiva).

Importou-se da base da Berkeley FrameNet o frame Competição (atividade esportiva regida por regras), tornando-o subframe de Cenário_do_esporte. A partir da competição, refletiu-se sobre as etapas que marcam esse evento, elaborou-se dois frames filhos por herança: o frame Fase_preliminar (fases em que um determinado número de competidores que obtiver os melhores resultados passa de fase para o evento oficial) e Torneio_de_eliminação (fases de um evento esportivo em que os atletas ou equipes competem entre si durante uma partida. O ganhador passa para a fase seguinte e o perdedor que fica é eliminado). Devido à relevância e importância da fase final de uma competição, o frame Torneio_de_eliminação impulsiona a origem do frame Final (a competição chega ao fim, com um atleta ou mais vencedores ganhando de um ou mais perdedores) com uma relação de precedência entre eles. Há uma relação de precedência (*Precedes*) marcada pelas setas pretas entre Fase_preliminar e Torneio_de_eliminação, e também entre Final e Cerimônias. O frame Cerimônias marca os eventos de abertura, encerramento e premiação dos atletas durante uma competição ou um evento esportivo como as Olimpíadas.

4.2.2 Relações

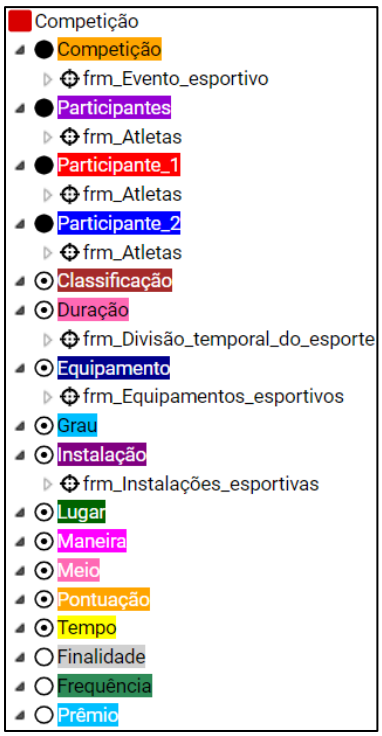
Como consequência desta pesquisa, a partir do ilustrado na subseção 3.2.2 com as estruturas de representação semântica postuladas por Pustejovsky e suas equivalências na base de dados da FrameNet, três novos tipos de relações foram implementadas e modeladas na base de dados da FrameNet Brasil: a relação EF a Frame, a relação de equivalência de tradução entre ULs e as relações qualia ternárias, sendo todas essas relações peculiares à FrameNet Brasil.

A relação EF-Frame trata de uma referência que o EF de um dado frame pode fazer a outro(s) frame(s), ou seja, um EF de um frame mapeado a outro frame sugere que os elementos que podem ocupar esse papel ilustrado no EF estão definidos e modelados no outro frame. Matos (2014) expõe que o objetivo primário dessa relação

> é estender a interpretação conceitual dos Elementos de Frame, mostrando que, além de atuarem em termos linguísticos como papéis microtemáticos, específicos para a situação descrita pelo Frame, eles podem desempenhar uma função cognitiva, como evocadores de outros Frames, a partir do Frame onde se situam. (MATOS, 2014, p.83).

A partir das colocações de Matos (2014) acerca da proposição da relação EF-frame, vejamos um exemplo desse tipo de relação.

Figura 46 – Representação da Relação EF-Frame a partir dos EFs do frame `Competição`



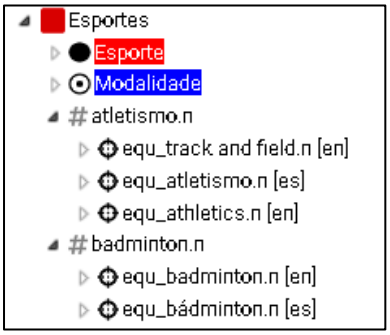
Fonte: Webtool da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

Na Figura 46, temos ilustrada essa primeira relação de EF a Frame dentro do frame `Competição`. Tomemos como exemplo o EF Participantes. É estabelecida uma relação entre esse EF e o frame `Atletas`. Normalmente um frame de evento, como o frame `Competição`, possui participantes que se engajam nesse evento. Esses participantes geralmente são entidades. No exemplo do frame `Competição`, o EF Participantes tende a se manifestar através das ULs definidas no frame `Atletas`. Tem-se, assim, uma relação do evento com as entidades que

tomam parte nele, contribuindo para o adensamento da base de dados. Ainda na Figura 46, pode-se constatar que os EFs Competição, Equipamento, Instalação e Duração têm suas manifestações na localidade sintática das ULs do frame `Competição` definidas nos termos dos frames `Evento_esportivo`, `Equipamentos_esportivos`, `Instalações_esportivas` e `Divisão_temporal_do_esporte`, respectivamente.[76]

O segundo tipo de relação implementada na FrameNet Brasil compreende as relações de equivalência de tradução estabelecidas entre as ULs dos domínios específicos do Turismo e dos Esportes. No âmbito desta tese, as relações de equivalência de tradução foram propostas para o domínio dos Esportes entre 1651 ULs do português, 2051 ULs do inglês e 1059 ULs do espanhol. Vejamos o exemplo das relações de equivalência nas ULs do frame `Esportes` ilustrado na Figura 47.

Figura 47 – Relações de Equivalência de Tradução nas ULs do frame `Esportes`.



Fonte: Webtool da FrameNet Brasil. (http://webtool.framenetbr.ufjf.br/).

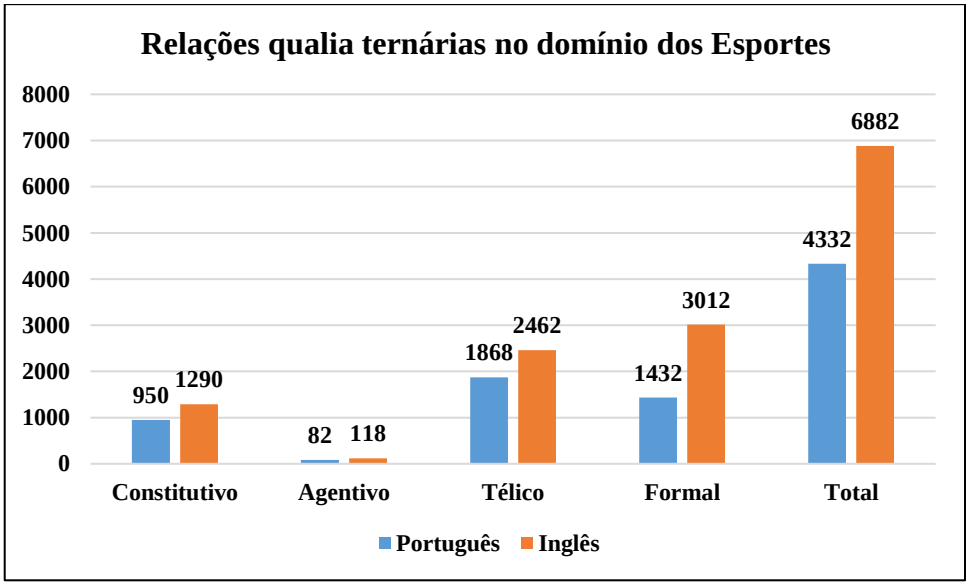
A Figura 47 traz as relações de equivalência de tradução para as ULs dos Esportes na FrameNet Brasil. Os equivalentes em língua inglesa para as ULs aparecem seguidos de [en]. Os equivalentes em espanhol das ULs se mostram seguidos de [es]. Por exemplo, o esporte **atletismo.n** em português possui os equivalentes de tradução ***track and field.n*** e ***athletics.n*** em inglês e ***atletismo.n*** em espanhol.

O último tipo de relação de caráter inovador criada na FrameNet Brasil são as relações qualia ternárias. Essas relações foram teoricamente apresentadas na subseção 3.2.2, ilustradas

[76] Uma vez que a FrameNet não lista como ULs as ENs, a princípio, poder-se-ia argumentar que, se o EF Participantes fosse manifestado em uma sentença através dos nomes dos atletas envolvidos, a relação de EF a frame seria inútil. Entretanto, como se mostrará mais adiante, a base de dados do m.knob também permite que os EFs e ULs sejam mapeados a classes em uma ontologia, o que poderia servir de ponto de conexão a bases de dados abertos ligados, adensando ainda mais a rede.

na Figura 37 e exemplificadas pela Figura 38. Para esta tese, o funcionamento e testes iniciais do tradutor do m.knob, foram modeladas relações qualia para o português e para o inglês. Analisemos a quantidade de relações qualia ternárias criadas a partir do gráfico.

Gráfico 1 – Relações qualia ternárias no domínio dos Esportes



Fonte: Elaborado pelo autor (2020).

Como se observa no Gráfico 1, das 4.332 relações qualia criadas para o português, 950 foram do tipo constitutivo, 82 do tipo agentivo, 1868 do tipo télico e 1.432 do tipo formal. Já para a língua inglesa, de um total de 6.882 relações criadas, 950 foram do tipo constitutivo, 118 do tipo agentivo, 2462 do tipo télico, 3.012 do tipo formal. A diferença entre o número de relações criadas em cada língua se deve à quantidade de ULs criadas em cada língua e às relações de equivalência de tradução estabelecida entre as ULs. Por exemplo, ao se estabelecer 4 relações qualia entre atletismo.n e outras ULs em português, em inglês o número de relações qualia será, no mínimo o dobro, dado o fato de atletismo.n possuir como equivalentes *athletics.n* e *track and field.n*.[77] Passemos à subseção 4.2.3 com o conceito de ontologias e seu uso na base de dados da FrameNet Brasil e no m.knob.

[77] As relações qualia ternárias foram criadas manualmente em português e foram replicadas computacionalmente para as ULs do inglês e do espanhol a partir das relações de equivalência de tradução que já haviam sido estabelecidas entre ULs do português, inglês e espanhol.

4.2.3 Ontologias

Outro recurso fundamental para a execução desta pesquisa são as ontologias. As ontologias são essenciais para que haja uma representação formal do conhecimento legível por computadores. Segundo Gruber (2009), "[...] uma ontologia define um conjunto de primitivos utilizados para se modelar um domínio de conhecimento ou discurso" (GRUBER, 2009, p. 1963)[78]. Com esses primitivos, o autor se refere a classes (ou conjuntos), atributos (ou propriedades) e algum tipo de relacionamento entre os membros das classes (relações). Esses primitivos são descritos conforme seus significados e restrições, que podem regular sua existência. Para o autor, as ontologias simbolizam um nível "semântico" de representação.

Helbig (2006) possui uma visão semelhante acerca de ontologia ao apresentá-la como

> [...] uma classificação de conceitos de um ponto de vista epistêmico, que até certo ponto também espelha uma classificação do mundo real segundo aspectos ontológicos. As classes de conceitos dessa ontologia são chamadas tipos. Esses tipos desempenham um papel importante ao estabelecerem um aparato formal de representação do significado. (HELBIG, 2006, p. 22)[79].

Para ambos os autores, uma ontologia é vista como uma representação formal do conhecimento através de categorias, elementos que preenchem essas categorias e relações entre elas. As ontologias contribuem para a formalização do conhecimento de certos domínios ou áreas de conhecimento, além de facilitarem o compartilhamento e a reutilização de informações.

O Sistema de Recomendação do m.knob conta com uma estruturação ontológica contida no Google Places. Segundo a documentação do Google Places[80], um local pode ser definido como um espaço físico contendo um nome ou algo que pode ser encontrado em um mapa. Como locais, citemos empresas, pontos de interesse e pontos geográficos específicos.

Conforme dados extraídos da documentação, o Google Places

> [...] fornece ao seu aplicativo informações valiosas sobre locais, incluindo o nome e o endereço do local, a localização geográfica especificada como coordenadas de latitude / longitude, o tipo de local (como casa noturna, loja de animais, museu) e muito mais. Para acessar essas informações de um local

[78] "[…] an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse." (GRUBER, 2009, p. 1963).
[79] "[…] A classification of concepts from an epistemic point of view, which to a certain degree also mirrors a classification of the real world according to ontological aspects. The classes of concepts defined by this ontology are also named sorts. Sorts play an important role in designing the formal apparatus of meaning representation." (HELBIG, 2006, p. 22).
[80] https://developers.google.com/places/android-api/start Acesso em: 01 jul. 2020.

> específico, você pode usar o ID do local, um identificador estável que identifica exclusivamente um local. (GOOGLE, 2020)[81].

O Sistema de Recomendação do m.knob retorna suas recomendações turísticas a partir da extração de locais do Google Places, selecionando o local de uma coleção de tipos (*regions*) e retornando quaisquer resultados relacionados a locais turísticos (ex.: bares, restaurantes, parques, igrejas etc.).

Com o propósito de adensar cada vez mais a base de dados, fora a modelagem de frames e ULs do domínio dos Esportes com base nos manuais de esportes, sites de associações esportivas e sites de notícias esportivas, recorremos também à extração automática de ULs de uma outra base de dados ligados abertos chamada BabelNet (NAVIGLI; PONZETTO, 2012)[82]. Após a extração, essas ULs foram validadas pelos linguistas da FrameNet Brasil. A BabelNet pode ser compreendida como um recurso computacional lexical que liga dados e informações extraídos da Wikipédia[83] com um dos mais populares *lexicons* da Web, a *WordNet*. Segundo Navigli e Ponzetto (2012), a BabelNet tem como resultado "[...] um 'dicionário enciclopédico' que fornece *babel synsets*, ou seja, conceitos e entidades nomeadas lexicalizados em muitas línguas e conectados através de uma grande quantidade de relações semânticas". (NAVIGLI; PONZETTO, 2012, p. 218)[84]. A função de tal recurso é fornecer uma cobertura lexicográfica e enciclopédica completa, incluindo 14 milhões de entradas e 271 línguas. Construindo uma ampla rede semântica, a BabelNet enfatiza os sentidos das palavras (conceitos) e as Entidades Nomeadas, conectando-os através de relações semânticas. O m.knob trabalha com o *bootstrapping* ou a extração das informações contidas nos *babel synsets*, havendo possíveis candidatos a equivalentes de tradução em cada *synset* para a palavra procurada. Foram extraídas para a base de dados da FrameNet 426 ULs em inglês e 190 ULs em espanhol dentro dos domínios do Turismo e dos Esportes. A partir dessa extração, o projeto FrameNet Brasil através de seus linguistas realizou uma validação lexicográfica dos possíveis equivalentes de tradução entre as ULs do português e essas modeladas na base de dados em inglês e espanhol (via relação de equivalência anteriormente ilustrada).

---

[81] "[…] The Places SDK for Android provides your app with rich information about places, including the place's name and address, the geographical location specified as latitude/longitude coordinates, the type of place (such as night club, pet store, museum), and more. To access this information for a specific place, you can use the place ID, a stable identifier that uniquely identifies a place." (GOOGLE, 2020).

[82] https://babelnet.org/

[83] https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:P%C3%A1gina_principal

[84] "[…] The result is an 'encyclopedic dictionary' that provides babel synsets, i.e., concepts and named entities lexicalized in many languages and connected with large amounts of semantic relations." (NAVIGLI; PONZETTO, 2012, p. 218).

5. METODOLOGIA EXPERIMENTAL

O Capítulo 5 expõe a metodologia utilizada nos testes de avaliação dos sistemas de tradução aqui propostos, passando pela constituição do corpus de sentenças de teste, o estabelecimento de traduções humanas de referência (*Gold Standard*), a avaliação sintático-semântica das traduções de referência, e por fim, a realização de uma avaliação do sistema de TM com injeção terminológica em pré-processamento (S-Pré) e do sistema de TM com injeção terminológica na pós-edição (S-Pós) através de métricas de avaliação mundialmente utilizadas, BLEU, TER e HTER. Será avaliado, ainda, o desempenho de um sistema de TM por RNN que representa o estado da arte implementado em um sistema comercial, qual seja o Google Tradutor (S-Base).

5.1 CORPUS FONTE DAS SENTENÇAS DE TESTE

Inicialmente, para a realização dos testes de tradução (do DAISY, do S-Pré e do S-Pós), foi constituído um corpus de 50 sentenças específicas do domínio dos esportes. Para que os testes se sucedessem de forma adequada e atingissem os objetivos esperados, foram determinadas algumas condições para as sentenças, tais como: (i) em cada sentença deveria haver pelo menos duas palavras dos esportes que fossem também ULs na base de dados da FN-Br, estivessem ligadas com pelo menos um tipo de relação qualia, não descartando a possibilidade de haver mais ULs ligadas via qualia dentro da mesma sentença; (ii) cada sentença deveria possuir uma palavra que fosse polissêmica, tendo um de seus significados nos esportes e pelo menos um outro significado em outro domínio, estando essa palavra do esporte modelada na base de dados; por último, (iii) as sentenças deveriam ser reais, ou seja, produzidas em contextos reais da língua e relacionados aos esportes.

Atendendo a esses critérios, foram compiladas as 50 sentenças específicas dos esportes em português a partir de sites, manuais, blogs e revistas. Essa compilação das sentenças de teste consta da Tabela 3.

Tabela 3 – Sentenças utilizadas nos testes do DAISY e dos Sistemas Pré e Pós de TM

Número	Sentença	Fonte
1	Um corredor não tenta correr uma maratona nos primeiros dias de treinos.	https://logosconcursos.com.br/noticias/os-5-primeiros-passos-de-um-concurseiro Acesso em: 25 set. 2019.
2	O lançador é desclassificado se sair da zona de lançamento antes, durante ou depois do lançamento.	https://www.olimpiadatododia.com.br/lima-2019/atletismo/lancamento-de-dardo-masculino/ Acesso em: 25 set. 2019.

3	O árbitro Mário Yamasaki decidiu interromper a luta achando que o lutador havia desmaiado.	https://nocautenarede.com.br/pos-lutas-ufc-fight-night-112-resumo-resultados-e-bonus/ Acesso em: 25 set. 2019.
4	Coloco o ponto em que o levantador executa o levantamento.	http://www.justvolleyball.com.br/vietart26_estrategia_ofensiva_bolasdetempo_introducao_cabecafrente.html Acesso em: 25 set. 2019.
5	O ponta é o jogador que menos tempo tem para pensar na armação de uma jogada.	Revista Placar n. 713 de 20/01/1984 - p. 41
6	O ginásio possui uma quadra que pode receber jogos de futsal e handebol.	http://suzano.sp.gov.br/web/wp-content/uploads/2019/08/volume%20II%20-%20Diagnostico%20Tur%C3%ADstico%20Suzano.docx Acesso em: 25 set. 2019.
7	O OG Kyle Long sofreu uma lesão na mão durante o primeiro quarto do jogo contra os Saints.	http://nfldebolsa.com/lista-de-lesionados-da-semana-8/ Acesso em: 25 set. 2019.
8	Além disso, a rede possui 1,55 cm sendo mais alta que a utilizada no tênis e menor que a rede da quadra de vôlei.	Revista Mês Ano 2 n. 16 - maio de 2012 – p. 51
9	A competição de saltos em equipe envolve um saltador e uma saltadora.	https://cbda.org.br/_uploads/saltos/RegrasOficiaisSaltosOrnamentais2017_2021.pdf Acesso em: 25 set. 2019.
10	O tênis, um dos esportes mais tradicionais e praticados no mundo.	https://clubepauloafonso.com.br/torneio-de-tenis/ Acesso em: 25 set. 2019.
11	A bandeja é quando o jogador faz a cesta bem próxima do aro.	http://globoesporte.globo.com/sc/especial-publicitario/federacao-catarinense-de-basketball/basquete-inspira-sc/noticia/2017/06/basquete-de-z.html#:~:text=A%20bandeja%20%C3%A9%20quando%20o,e%20a%20segunda%20mais%20longa.&text=Charge%3A%20Em%20tradu%C3%A7%C3%A3o%20livre%2C%20pode,ser%20chamada%20de%20%E2%80%9Ccarga%E2%80%9D. Acesso em: 25 set. 2019.
12	Durante uma jogada, um jogador pode dar até dois toques não consecutivos, de modo que a equipe só pode dar no total três toques na bola.	https://docer.pl/doc/sxe5c00 Acesso em: 25 set. 2019.
13	Mario Suárez chuta rente à trave do gol de Schwarzer.	http://www.espn.com.br/video/405392_tempo-real-quase-mario-suarez-chuta-rente-a-trave-do-gol-de-schwarzer Acesso em: 25 set. 2019.
14	A vela é um esporte olímpico desde 1900.	http://fnb.org.br/iatismo-muito-mais-que-um-esporte/ Acesso em: 25 set. 2019.
15	No último lance do jogo, o zagueiro Gustavo Gómez deu um carrinho no atacante corinthiano Jô e fez o pênalti.	https://www.esporteinterativo.com.br/futebolbrasileiro/Torcida-do-Palmeiras-massacra-Gustavo-Gomez-nas-redes-sociais-20200808-0036.html Acesso em: 25 set. 2019.
16	O artilheiro é o jogador Evandro, que marcou sete gols.	http://www.gritoregional.com.br/mm-direcoes-e-flamenguinho-decidem-hoje-tarde-o-12o-campeonato-de-futebol-suico-na-cohab/ Acesso em: 25 set. 2019.
17	Fuga é uma técnica utilizada no ciclismo de estrada.	https://pt.wikipedia.org/wiki/Fuga_(ciclismo)#:~:text=Fuga%20%C3%A9%20uma%20t%C3%A9cnica%20utilizada,e%20favorecer%20o%20sprint%20final. Acesso em: 25 set. 2019.
18	A partida se inicia sempre com um saque, jogada que, obrigatoriamente, alterna-se entre os participantes a cada game.	https://www.intrinseca.com.br/blog/2019/11/a-maior-batalha-fisica-e-mental-do-esporte-aprenda-tudo-sobre-o-tenis/ Acesso em: 25 set. 2019.
19	Jogando na posição 3, um ala é o jogador que mais se aproxima dos dois extremos das posições do basquete.	https://www.theplayoffs.com.br/blog/blog-nba/entenda-jogo-conheca-tradicionais-posicoes-jogadores-basquete/ Acesso em: 25 set. 2019.
20	A equipe jogava bem e Juciely, com uma china, jogada muito utilizada pela atleta, fez 09-06.	https://www.tabelacarioca.com.br/2020/03/06/sesc-rj-faz-jogo-tranquilo-e-vence-o-sao-caetano-em-casa/ Acesso em: 06 mar. 2020.
21	No jogo de volta da decisão, torcedores do River apedrejaram o ônibus dos jogadores do Boca Juniors, que não tiveram condições de entrar no campo do Estádio Monumental de Nuñez, em Buenos Aires.	https://veja.abril.com.br/esporte/final-da-libertadores-entre-flamengo-e-river-sera-em-lima/#:~:text=A%20decis%C3%A3o%20da%20Conmebol%20vem,de%20Nu%C3%B1ez%2C%20em%20Buenos%20Aires. Acesso em: 05 nov. 2019.
22	Com 18 anos, o capitão do time foi o jogador mais jovem da história do clube a marcar em um Atletiba.	https://globoplay.globo.com/v/3673724/ Acesso em: 25 set. 2019.

23	Então a piscina tem de ser dividida em dez raias para que só as oito internas, menos turbulentas, sejam usadas nas provas.	https://super.abril.com.br/mundo-estranho/como-e-uma-piscina-olimpica/
24	O teoricamente dono da posição está novamente sem atuar, desta vez por lesão, enquanto o reserva é o jogador que mais vezes atuou nesta temporada.	http://www.espn.com.br/noticia/364895_em-boa-fase-aloisio-deixa-interrogacao-no-ataque-para-muricy Acesso em: 25 set. 2019.
25	O cruzamento é uma jogada forte nossa, como é de todas as equipes.	https://www.correiodopovo.com.br/esportes/gr%C3%AAmio/par%C3%A1-retruca-dami%C3%A3o-e-diz-que-n%C3%A3o-tem-nada-de-final-antecipada-1.90371 Acesso em: 25 set. 2019.
26	No entanto, a mídia esportiva costuma referir à jogada apenas como bicicleta, pouco empregando o prefixo chute ou pontapé.	https://pt.wikipedia.org/wiki/Bicicleta_(futebol)#:~:text=No%20entanto%2C%20a%20m%C3%ADdia%20esportiva,chute%22%20ou%20%22pontap%C3%A9%22.&text=Na%20l%C3%ADngua%20inglesa%2C%20seu%20nome,executor%20estar%20pedalando%20uma%20bicicleta. Acesso em: 25 set. 2019.
27	O estádio possui uma pista de atletismo de nove raias, dois telões gigantes e uma rede wi-fi de última geração.	http://www.cefsa.org.br/home/sobre-o-cefsa/estadio-olimpico/ Acesso em: 25 set. 2019.
28	Quatro países irão disputar o desafio dos quatro estilos da natação, onde borboleta é o nado mais complexo de aprender.	https://globoplay.globo.com/v/2152275/ Acesso em: 25 set. 2019.
29	Segundo dados do Footstats, o ponta foi o jogador com mais cruzamentos certos e mais passes para gol na equipe.	http://blogs.lance.com.br/numeros-da-bola/na-mira-do-vasco-kelvin-liderou-duas-estatisticas-do-sao-paulo-no-brasileiro/ Acesso em: 25 set. 2019.
30	Totalmente diferente dos outros estilos, o nado peito exige muita coordenação e técnica do praticante.	https://www.ativo.com/natacao/treinamento-natacao/tecnicas-de-natacao-nado-peito/ Acesso em: 25 set. 2019.
31	Pedrinho dá um chapéu, mas não dá sequência na jogada, aos doze minutos do primeiro tempo.	http://globoesporte.globo.com/tempo-real/videos/v/pedrinho-da-um-chapeu-mas-nao-da-sequencia-na-jogada-aos-12-do-1o-tempo/8112855/ Acesso em: 25 set. 2019.
32	A tradicional comemoração com um peixinho na quadra está viva na memória.	A tradicional comemoração com um peixinho na quadra está viva na memória. Acesso em: 25 set. 2019.
33	A posição e o alinhamento da gaiola no campo de competição é, portanto, crítico para o seu uso seguro.	https://www.passeidireto.com/arquivo/56333508/regras-oficiais-2018-2019 Acesso em: 25 set. 2019.
34	O nado de costas causa uma boa sensação após a execução de séries intensas de crawl, ou livre, e borboleta.	https://www.ativo.com/natacao/noticias-natacao/tecnicas-de-natacao-nado-de-costas-para-iniciantes/#:~:text=Muito%20utilizado%20pelos%20nadadores%20em,por%20nadadores%20especialistas%20nesse%20estilo. Acesso em: 25 set. 2019.
35	Muitos diziam que este era um salto criado pela escola queniana de atletismo, mas me parece que é uma variante do salto tesoura.	https://www.mdig.com.br/index.php?itemid=38859 Acesso em: 25 set. 2019.
36	Muitos acham que o gancho é o golpe onde o lutador lança sua mão de baixo para cima.	http://guerreirotkd.blogspot.com/2012/07/origem-e-regras-do-boxe.html Acesso em: 25 set. 2019.
37	Companheiro de Messi no Barcelona, o meia é o jogador brasileiro com o maior valor a atuar na Copa América.	https://www.gazetadopovo.com.br/esportes/top-10-jogadores-mais-valiosos-da-copa-america-2019/#:~:text=Companheiro%20de%20Messi%20no%20Barcelona,(R%24%20393%20milh%C3%B5es).&text=O%20meia%2Datacante%20da%20Juventus,(R%24%20372%20milh%C3%B5es). Acesso em: 25 set. 2019.
38	Cinco séries de golpes combinados de suple, bombeiro e estabilização no solo com ênfase na precisão de movimento.	https://pt.calameo.com/books/003702468a3838cc60248 Acesso em: 25 set. 2019.
39	No rugby, o abertura é o jogador mais habilidoso do time.	https://rugbybrasilrs.wordpress.com/2012/12/13/manual-basico-do-rugby-posicoes/#:~:text=Abertura%20%E2%80%93%20N%C2%BA%2010%20%3A%20Tal%20qual,comandar%20as%20a%C3%A7%C3%B5es%20dos%20backs. Acesso em: 25 set. 2019.
40	O servidor é o jogador que coloca a bola em jogo para o primeiro ponto.	https://docplayer.com.br/amp/161237578-Pan-american-masters-games-rio-2020-felipe-toledo-diretor-tecnico-abterj.html Acesso em: 25 set. 2019.

41	O sonho de Tristan Garcia, de 14 anos e fã de basquetebol era arremessar uma bola na cesta da quadra da escola.	https://www.sonoticiaboa.com.br/2019/06/12/estudantes-ajudam-garoto-com-paralisia-fazer-1-cesta-vida-assista/ Acesso em: 25 set. 2019.
42	Um arco é um equipamento individual e pessoal.	http://bowarrowtech.blogspot.com/2013/10/guia-pratico-para-dimensionamento-de.html Acesso em: 25 set. 2019.
43	Invista em uma boa luva, a luva é o equipamento de segurança maior do boxe, ela pode evitar que você se machuque gravemente.	https://mcz10.com/estilo-de-vida/esquadrao-bem-estar/o-boxe-e-um-aliado-peso-pesado-a-sua-saude/#:~:text=10%2D%20Invista%20em%20uma%20boa,e%20escoria%C3%A7%C3%B5es%20como%20%E2%80%9Cesfolamento%E2%80%9D. Acesso em: 25 set. 2019.
44	Todos os ginastas que disputam a prova saltam sobre um aparelho ligeiramente inclinado chamado mesa.	https://pt.wikipedia.org/wiki/Mesa_(gin%C3%A1stica) Acesso em: 25 set. 2019.
45	Se a ginástica rítmica é a ovelha negra da família ginástica, em seguida, a corda é a ovelha negra dos aparelhos.	https://ginasticaritmicabrasil.wordpress.com/2015/12/15/ginastica-ritmica-seria-a-corda-a-ovelha-negra-da-gr/ Acesso em: 25 set. 2019.
46	A fita é considerada o aparelho mais plástico e característico da ginástica.	https://pt.wikipedia.org/wiki/Fita_(gin%C3%A1stica)#:~:text=A%20fita%20%C3%A9%20considerada%20o%20aparelho%20mais%20pl%C3%A1stico%20e%20caracter%C3%ADstico%20da%20gin%C3%A1stica. Acesso em: 25 set. 2019.
47	Ela venceu as disputas nos aparelhos: fita e maças.	http://www.gaz.com.br/conteudos/esportes/2019/10/21/156315-alice_silva_e_campea_brasileira_em_torneio_de_ginastica_ritmica.html.php#:~:text=Ela%20venceu%20as%20disputas%20nos,Bernardo%20do%20Campo%20(SP). Acesso em: 25 set. 2019.
48	A vara para salto é um equipamento muito avançado.	http://fisicadesporto.blogspot.com/2011/03/vara-para-o-salto-vara.html#:~:text=A%20vara%20para%20salto%20%C3%A9,quando%20volta%20%C3%A0%20posi%C3%A7%C3%A3o%20normal. Acesso em: 25 set. 2019.
49	O atacante ainda protagonizou um lance de brilho ao aplicar um lençol em um adversário, jogada que chamou atenção de outros boleiros na web.	https://www.bol.uol.com.br/listas/lencol-de-neymar-e-coutinho-vaiado-marcam-rodada-de-brasileiros-na-europa.htm Acesso em: 25 set. 2019.
50	A prancha é o equipamento mais importante para pegar ondas grandes.	https://www.mormaii.com.br/site/dicas-pra-surfar-ondas-grandes-com-carlos-burle/#:~:text=A%20prancha%20%C3%A9%20o%20equipamento,prancha%20e%20de%20boa%20qualidade. Acesso em: 25 set. 2019.

Fonte: Elaborado pelo autor (2020).

Com o propósito de ilustrar a polissemia existente nos termos específicos dos esportes contidos nas 50 sentenças selecionadas, analisemos a Tabela 4 com as definições para os termos no domínio dos esportes, domínio genérico e em outros domínios.

Tabela 4 – Termos polissêmicos dos Esportes, suas definições e outras acepções

Sentença	**Termo**	**Significado**
1	corredor.n	Esportes: atleta que ou o que participa de corridas a pé, a cavalo, de automóvel, de moto etc.
		Arquitetura: no interior de construção, passagem que serve de ligação entre um ou mais cômodos.
2	lançador.n	Esportes: atleta que lança, arremessa ou atira em algum esporte.
		Astronáutica: profissional que se destina a conduzir ao espaço uma lançadeira (diz-se de foguete).
		Economia: aquele que vende opção ('documento negociável') nas bolsas de valores.

2	zona de lançamento.n	Esportes: no atletismo, trata-se da área na qual o atleta é permitido ficar durante provas de lançamentos de disco, peso, dardo ou martelo.
		Astronáutica: estrutura especialmente preparada para o lançamento de veículos espaciais, como foguetes e ônibus espaciais
2	lançamento.n	Esportes: arremesso, jogada, passe.
		Publicidade: o período inicial de uma campanha publicitária.
		Jurídico: procedimento por meio do qual o juiz declara achar-se expirado o prazo para apresentação de provas ou documentos em uma ação.
3	árbitro.n	Esportes: aquele que faz cumprir, numa competição ou disputa, as regras estabelecidas para a modalidade de esporte que está sendo praticada; juiz.
		Jurídico: aquele que, por acordo das partes interessadas ou designação de tribunal, é indicado para dirimir uma questão; mediador, juiz.
3	luta.n	Esportes: combate, esp. de caráter esportivo, em que dois adversários desarmados se enfrentam em corpo a corpo.
		Genérico: esforço para superar, para vencer obstáculos ou dificuldades.
3	lutador.n	Esportes: atleta que luta profissionalmente em algum esporte.
	lutador.a	Genérico: aquele que é dotado de espírito de luta, que está sempre pronto a defender alguém ou uma causa.
4	ponto.n	Esportes: cada uma das unidades que, em competições diversas, marcam as posições dos competidores com relação a seus ganhos ou perdas.
		Geometria: intersecção de duas retas; conceito primitivo da geometria que representa uma figura geométrica sem dimensões.
		Costura/Alfaiataria: pequeno orifício feito com agulha que se enfia em tecido, couro etc., para fazer passar o fio.
4	levantador.n	Esportes: aquele que, em determinados jogos, como o vôlei, p.ex., tem a função de levantar a bola para que outro jogador a golpeie.
		Cirurgia: diz-se de ou instrumento com que se levantam do cérebro os fragmentos dos ossos fraturados do crânio.
4	levantamento.n	Esportes: no vôlei, o lançamento da bola ao alto, ger. perto da rede, para que outro jogador possa golpeá-la com força na direção do time adversário.
		Estatística: conjunto de operações para determinar o número de ocorrências, as intensidades ou as modalidades dos fenômenos individuais que compõem um ou mais fenômenos coletivos.
5	ponta.n	Esportes: atacante que joga preferencialmente numa das pontas (acp. 22); ponteiro.
		Genérico: parte extrema de um objeto, considerado longitudinalmente; extremidade.
5	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
6	ginásio.n	Esportes: sala, estabelecimento destinado à prática da cultura física ou de esportes, freq. com acomodações para plateia.
		Educação: escola, estabelecimento onde é ministrado esse curso.
6	quadra.n	Esportes: área retangular demarcada de forma a permitir a prática de determinados esportes.
		Genérico: distância de uma esquina a outra, no mesmo lado da rua.
6	jogo.n	Esportes: atividade para entreter regida por regras em que normalmente um jogador perde e o outro ganha.
		Engenharia mecânica: mecanismo de direção de um veículo.
7	quarto.n	Genérico: aposento ou divisão da casa onde se dorme.
		Esportes: medida de tempo que pode se referir a um dos tempos do basquete.
7	jogo.n	Esportes: atividade para entreter regida por regras em que normalmente um jogador perde e o outro ganha.
		Engenharia mecânica: mecanismo de direção de um veículo.
8	rede.n	Costura/Alfaiataria: entrelaçado de fios, de espessura e materiais diversos, formando um tecido de malhas com espaçamentos regulares.
		Informática: sistema constituído pela interligação de dois ou mais computadores e seus periféricos, com o objetivo de comunicação, compartilhamento e intercâmbio de dados.
		Esportes: equipamento feito de rede, que se estende no centro da quadra de tênis, voleibol etc., sobre o qual a bola deve passar para continuar em jogo.
8	quadra.n	Esportes: área retangular demarcada de forma a permitir a prática de determinados esportes.
		Genérico: distância de uma esquina a outra, no mesmo lado da rua.
9	competição.n	Ecologia: interação intra ou interespecífica que ocorre quando duas ou mais espécies necessitam de um mesmo recurso ambiental limitado.
		Esportes: prova que põe em concorrência duas ou mais pessoas ou grupos no que tange a determinadas aptidões ou qualidades físicas ou atléticas.

9	salto.n	Vestuário: tacão de calçado.
		Esportes: ação ou efeito de saltar; pulo.
9	saltador.n	Ornitologia/Aves: Pássaro TIZIO (Volatinia jacarina).
		Esportes: atleta especializado em provas de salto.
10	tênis.n	Vestuário/Chapelaria: sapato de material leve (lona, tecido, couro, plástico) e sola flexível de borracha, para uso esportivo e geral; sapato-tênis.
		Esportes: modalidade de esporte em que dois ou quatro jogadores (no caso de duplas), munidos de raquete, impelem uma bola especial por cima de uma rede que divide a quadra em dois campos.
11	bandeja.n	Culinária: recipiente raso us. para o serviço de alimentos, bebidas etc. ou como peça decorativa.
		Esportes: no basquete, jogada em que o atleta encesta a bola, conduzindo-a por baixo, com uma das mãos.
11	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
11	cesta.n	Artesanato: utensílio próprio para a guarda de objetos diversos, feito de fibras entrançadas, provido ou não de alças e/ou tampa, conforme o uso; cesto.
		Esportes: aro de metal fixado à tabela ('suporte retangular') que sustenta uma malha sem fundo onde a bola é arremessada durante o jogo; cesto.
12	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
12	toque.n	Medicina: exame de uma cavidade corporal praticado com o auxílio dos dedos.
		Esportes: ação ou efeito de tocar; tocamento, contato.
13	trave.n	Construção Civil: grande tronco ou madeiro retilíneo, grosso e comprido, us. para sustentar partes elevadas de uma construção.
		Esportes: os postes laterais do gol; o gol inteiro, inclusive o travessão.
14	vela.n	Genérico: peça ger. cilíndrica, de cera ou outra substância gordurosa, com um pavio central em toda a extensão, e cuja chama serve para iluminar.
		Esportes: peça de tecido de linho, algodão ou náilon us. para propulsão eólica de embarcação; pano. Esporte com barcos movidos à vela.
15	lance.n	Pesca: ação de pescar com rede.
		Esportes: movimento, jogada.
15	jogo.n	Esportes: atividade para entreter regida por regras em que normalmente um jogador perde e o outro ganha.
		Engenharia mecânica: mecanismo de direção de um veículo.
15	carrinho.n	Genérico: carro para transportar crianças pequenas. Carro de brinquedo.
		Esportes: lance em que um jogador procura retirar a bola de um adversário, atirando-se a seus pés com as pernas estendidas para a frente e deslizando sentado pelo chão.
15	atacante.n	Genérico: o que ataca; agressor.
		Esportes: jogador(a) que joga no ataque; dianteiro, avante.
16	artilheiro.n	Militar: militar (oficial ou soldado) pertencente à arma de artilharia.
		Esportes: em futebol, polo aquático, handebol e outros esportes, jogador que faz gol.
16	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
16	marcar.v	Genérico: identificar com marca, etiqueta, número etc.
		Esportes: apontar, registrar, fazer gol, ponto, atentar-se a outro jogador.
17	fuga.n	Genérico: ato ou efeito de fugir.
		Esportes: no ciclismo, é quando um ciclista aumenta o ritmo para escapar de um pelotão.
18	partida.n	Genérico: ato de partir; saída.
		Esportes: peleja esportiva; jogo, prélio.
18	saque.n	Economia: expedição de título de crédito ou ordem de pagamento em favor de si próprio ou de outrem.
		Esportes: no tênis, no voleibol etc., ação de pôr a bola em jogo, lançando-a por cima da rede em direção ao campo adversário; serviço.
19	ala.n	Engenharia: cada um dos resguardos laterais de uma ponte.
		Esportes: cada uma das porções laterais do ataque de certos desportos de equipe, como basquete e futebol.
19	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
20	china.n	Agropecuária: certa raça bovina.
		Esportes: no vôlei, é a jogada executada pela lateral, saída de rede. Ocorre quando o atacante se desloca de forma que salte somente em um dos pés.
21	jogo.n	Esportes: atividade para entreter regida por regras em que normalmente um jogador perde e o outro ganha.

		Engenharia mecânica: mecanismo de direção de um veículo.
21	decisão.n	Genérico: ato ou efeito de decidir; determinação.
		Esportes: decisões que o árbitro toma durante o jogo.
21	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
21	campo.n	Agropecuária: terreno plano e extenso destinado à agricultura ou às pastagens.
		Esportes: lugar próprio para a prática de diversos esportes.
22	capitão.n	Militar: comandante de número expressivo de combatentes.
		Esportes: jogador que comanda o time e fala pelos jogadores.
22	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
22	marcar.v	Genérico: identificar com marca, etiqueta, número etc.
		Esportes: apontar, registrar, fazer gol, ponto, atentar-se a outro jogador.
23	raia.n	Ictiologia/Peixes: design. comum aos peixes elasmobrânquios da ordem dos rajiformes, que ger. possuem corpo discoidal com nadadeiras peitorais muito desenvolvidas, cinco pares de fendas branquiais na região ventral, cauda com ou sem ferrão, e são bentônicos e ovovivíparos; arraia.
		Esportes: linha que delimita o espaço em que dois atletas competem, sendo numa pista de corrida ou em esportes aquáticos como a natação ou a canoagem.
23	prova.n	Jurídico: fato, circunstância, indício, testemunho etc., que demonstram a culpa ou a inocência de um acusado.
		Esportes: competição esportiva, ger. individual.
24	reserva.n	Jurírico: cláusula de contrato, escritura, etc., que limita, em qualquer aspecto, os seus efeitos.
		Esportes: atleta que substitui o titular de uma equipe, quando necessário; suplente.
		Militar: conjunto dos cidadãos que cumpriram os deveres militares, ou deles foram dispensados, e que se mantêm à disposição das forças armadas para casos de necessidade.
24	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
25	cruzamento.n	Biologia: acasalamento entre organismos distintos morfológica ou geneticamente.
		Esportes: jogada em que um jogador lança a bola para a posição em que o outro jogador se encontra.
26	bicicleta.n	Transporte: veículo composto de um quadro ('conjunto de tubos metálicos'), assentado sobre duas rodas iguais alinhadas uma atrás da outra e com raios metálicos, das quais a da frente é comandada por um guidom e funciona como diretriz, e a de trás, ligada a um sistema de pedais acionados pelo bicicletista, funciona como motriz.
		Esportes: lance em que um jogador, ger. atacante, de costas para o gol adversário, gira o corpo e chuta a bola por cima da cabeça, para trás; puxeta.
27	pista.n	Genérico: vestígio, rasto.
		Esportes: caminho, pavimento ou espaço especialmente preparado para a realização de corridas ou para a prática de exercícios.
27	raia.n	Ictiologia/Peixes: design. comum aos peixes elasmobrânquios da ordem dos rajiformes, que ger. possuem corpo discoidal com nadadeiras peitorais muito desenvolvidas, cinco pares de fendas branquiais na região ventral, cauda com ou sem ferrão, e são bentônicos e ovovivíparos; arraia.
		Esportes: linha que delimita o espaço em que dois atletas competem, sendo numa pista de corrida ou em esportes aquáticos como a natação ou a canoagem.
27	rede.n	Costura/Alfaiataria: entrelaçado de fios, de espessura e materiais diversos, formando um tecido de malhas com espaçamentos regulares.
		Informática: sistema constituído pela interligação de dois ou mais computadores e seus periféricos, com o objetivo de comunicação, compartilhamento e intercâmbio de dados.
		Esportes: equipamento feito de rede, que se estende no centro da quadra de tênis, voleibol etc., sobre o qual a bola deve passar para continuar em jogo.
28	estilo.n	Anatomia Zoológica: nos insetos, cerda espessa no ápice do terceiro segmento antenal.
		Esportes: tipo de movimento que um atleta escolhe ou tem que realizar em uma competição.
28	borboleta.n	Entomologia/Insetos: design. comum a todas as spp. de insetos lepidópteros da subordem dos ropalóceros; ger. diurnos, possuem antenas com as extremidades apicais dilatadas e asas sem frênulos; panapaná.

		Esportes: na natação, é o estilo com os ombros alinhados à superfície de água, o nadador faz o movimento simultâneo dos braços. As pernas alinhadas também vão para cima e para baixo simultaneamente.
29	ponta.n	Esportes: atacante que joga preferencialmente numa das pontas (acp. 22); ponteiro.
		Genérico: parte extrema de um objeto, considerado longitudinalmente; extremidade.
29	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
29	cruzamento.n	Biologia: acasalamento entre organismos distintos morfológica ou geneticamente.
		Esportes: jogada em que um jogador lança a bola para a posição em que o outro jogador se encontra.
29	passe.n	Religião: ato de passar as mãos repetidas vezes por diante ou por cima de pessoa que se pretende magnetizar ou curar pela força mediúnica.
		Esportes: ato ou efeito de entregar um equipamento esportivo, geralmente uma bola, para outro atleta da mesma equipe. No handball, é a ação de enviar e encaminhar a bola ao companheiro, de forma correta, para facilitar a próxima ação. No rugbi, é a passagem da bola com as mãos ou os pés.
30	estilo.n	Anatomia Zoológica: nos insetos, cerda espessa no ápice do terceiro segmento antenal.
		Esportes: tipo de movimento que um atleta escolhe ou tem que realizar em uma competição.
30	peito.n	Anatomia Geral: região do tronco que vai do pescoço ao abdome; tórax.
		Esportes: na natação, é o estilo mais lento, com o movimento simultâneo das mãos e dos pés.
31	chapéu.n	Vestuário/Chapelaria: peça do vestuário provida de copa e abas, destinada a cobrir a cabeça, ger. como adorno.
		Esportes: no futebol, é quando o jogador joga a bola por cima da cabeça do marcador e corre pelo lado em busca da bola.
31	tempo.n	Genérico: duração relativa das coisas que cria no ser humano a ideia de presente, passado e futuro; período contínuo no qual os eventos se sucedem.
		Esportes: cada um dos períodos em que se dividem as partidas de determinados jogos.
32	peixinho.n	Ictiologia/Peixes: peixe pequeno.
		Esportes: no vôlei, a mesma jogada na tentativa de evitar, com um passe, que a bola toque no chão.
32	quadra.n	Esportes: área retangular demarcada de forma a permitir a prática de determinados esportes.
		Genérico: distância de uma esquina a outra, no mesmo lado da rua.
33	gaiola.n	Genérico: caixa formada por um engradado de arame ou de ripas finas, destinada a aprisionar pássaros.
		Esportes: equipamento composto de um conjunto de três telas ou grades em forma de um quadrado, ficando a última parte aberta, em que o atleta pratica o lançamento de disco. Esse equipamento destina-se à proteção para que o disco não seja arremessado contra à arquibancada.
33	campo.n	Agropecuária: terreno plano e extenso destinado à agricultura ou às pastagens.
		Esportes: lugar próprio para a prática de diversos esportes.
33	competição.n	Ecologia: interação intra ou interespecífica que ocorre quando duas ou mais espécies necessitam de um mesmo recurso ambiental limitado.
		Esportes: prova que põe em concorrência duas ou mais pessoas ou grupos no que tange a determinadas aptidões ou qualidades físicas ou atléticas.
34	costas.n	Anatomia Geral: dorso ('região posterior').
		Esportes: na natação, é o estilo em que os nadadores nadam de costas.
34	livre.a	Genérico: que é senhor de si e de suas ações.
	livre.n	Esportes: na natação, é o movimento no qual o nadador, de barriga para baixo, dá três braçadas e vira a cabeça para o lado para respirar. Este estilo de nado também é conhecido como crawl ou crol e as pernas ficam balançando para cima e para baixo alternadamente.
34	borboleta.n	Entomologia/Insetos: design. comum a todas as spp. de insetos lepidópteros da subordem dos ropalóceros; ger. diurnos, possuem antenas com as extremidades apicais dilatadas e asas sem frênulos; panapaná.
		Esportes: na natação, é o estilo com os ombros alinhados à superfície de água, o nadador faz o movimento simultâneo dos braços. As pernas alinhadas também vão para cima e para baixo simultaneamente.
35	salto.n	Vestuário: tacão de calçado.
		Esportes: ação ou efeito de saltar; pulo.

35	tesoura.n	Genérico: utensílio para cortar, formado por duas lâminas de aço que se movem em cruz, unidas num eixo.
		Esportes: tipo de salto em que, ao saltar, a perna dianteira é erguida sobre a barra e, em seguida, a outra perna.
36	gancho.n	Genérico: haste recurva, de metal ou outra substância resistente, us. para suspender pesos ou pendurar objetos.
		Esportes: no boxe, é o golpe aplicado na linha da cintura do oponente.
36	golpe.n	Genérico: choque de um corpo com outro, que resulta em impacto (de pequena ou grande intensidade); pancada, batida.
		Esportes: recurso de ataque e defesa em luta corporal.
36	lutador.n	Esportes: atleta que luta profissionalmente em algum esporte.
	lutador.a	Genérico: aquele que é dotado de espírito de luta, que está sempre pronto a defender alguém ou uma causa.
37	meia.n	Vestuário/Chapelaria: peça de vestuário que calça os pés e alcança, de acordo com o modelo (soquete, três-quartos, comprida), diferentes alturas da perna ou da coxa.
		Esportes: no futebol, é a posição do atleta que atua no meio de campo entre a defesa e o ataque, criando jogadas ofensivas. No handebol, é a posição do atleta que é forte e muito importante nas ações ofensivas e defensivas, possuindo um forte arremesso a longa distância.
37	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
38	golpe.n	Genérico: choque de um corpo com outro, que resulta em impacto (de pequena ou grande intensidade); pancada, batida.
		Esportes: recurso de ataque e defesa em luta corporal.
38	bombeiro.n	Militar: membro de corporação que se destina a prestar socorro em casos de incêndio ou de sinistro.
		Esportes: na luta olímpica, é o golpe que consiste em o lutador segurar a coxa do oponente, puxando o braço e o lançando sobre os ombros.
38	estabilização.n	Genérico: ato ou efeito de estabilizar(-se).
		Esportes: em lutas, é o movimento em que um lutador consegue prender o outro por um determinado tempo.
38	solo.n	Pedologia: matéria orgânica ou mineral não consolidada aflorante, que mostra os efeitos de fatores genéticos e ambientais a que foi submetida.
		Esportes: estrado utilizado em provas de ginástica artística e lutas feito de um material elástico que amortece eventuais quedas e ajuda ao impulso dos saltos.
39	abertura.n	Rádio/TV: apresentação que inicia determinado programa de rádio ou TV, ger. padronizada para a série de programas.
		Esportes: no rúgbi, é a posição do atleta que orquestra o desempenho da equipe, recebe a bola e decide chutar, passar ou fazer uma pausa.
39	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
40	servidor.n	Informática: computador us. numa rede para proporcionar algum tipo de serviço (como acesso a arquivos ou a periféricos compartilhados) aos demais componentes da rede.
		Esportes: no tênis, badminton e tênis de mesa, é a posição do jogador que põe a bola em jogo.
40	jogador.n	Esportes: atleta que tem por profissão jogar.
		Genérico: aquele que tem o vício do jogo (de azar).
40	jogo.n	Esportes: atividade para entreter regida por regras em que normalmente um jogador perde e o outro ganha.
		Engenharia mecânica: mecanismo de direção de um veículo.
40	ponto.n	Esportes: cada uma das unidades que, em competições diversas, marcam as posições dos competidores com relação a seus ganhos ou perdas.
		Geometria: intersecção de duas retas; conceito primitivo da geometria que representa uma figura geométrica sem dimensões.
		Costura/Alfaiataria: pequeno orifício feito com agulha que se enfia em tecido, couro etc., para fazer passar o fio.
41	cesta.n	Artesanato: utensílio próprio para a guarda de objetos diversos, feito de fibras entrançadas, provido ou não de alças e/ou tampa, conforme o uso; cesto.
		Esportes: aro de metal fixado à tabela ('suporte retangular') que sustenta uma malha sem fundo onde a bola é arremessada durante o jogo; cesto.
41	quadra.n	Esportes: área retangular demarcada de forma a permitir a prática de determinados esportes.
		Genérico: distância de uma esquina a outra, no mesmo lado da rua.

42	arco.n	Arquitetura: forma arquitetônica ornamental, constituída por pilares ou colunas e pela estrutura curva (arco) ou semicurva que se encontra em sua parte superior externa ou interna.
		Esportes: no tiro com arco, é o equipamento impulsor que se usa para disparar flechas sobre qualquer alvo distante. Na ginástica rítmica, é o equipamento utilizado no formato de uma circunferência ou um cilindro fino de plástico ou madeira, também conhecido como bambolê.
43	luva.n	Serralheria: peça de ferro, plástico etc., provida ou não de rosca, us. para conexão de tubos e canos.
		Esportes: peça de vestuário utilizada para e cobrir e proteger as mãos, podendo ser utilizadas em alguns esportes pelos goleiros ou por lutadores de boxe ou jogadores de beisebol.
44	prova.n	Jurídico: fato, circunstância, indício, testemunho etc., que demonstram a culpa ou a inocência de um acusado.
		Esportes: competição esportiva, ger. individual.
44	aparelho.n	Anatomia Geral: união de dois ou mais sistemas.
		Esportes: na ginástica, é qualquer elemento utilizado para executar os movimentos e exercícios.
44	mesa.n	Mobília: móvel composto de um tampo horizontal, que ger. se destina a fins utilitários: refeições, jogos, escrita, costura, apoio etc.
		Esportes: na ginástica artística, é uma estrutura de metal coberta por uma textura almofadada e elástica para saltos.
45	corda.n	Genérico: feixe alongado de fibras vegetais (sisal, cânhamo etc.) ou matéria flexível similar, torcidas em espiral, de grossura e comprimento variáveis.
		Esportes: equipamento composto por um feixe de fibras trançadas ou enroladas entre si para a fixação de objetos ou a segurança de pessoas durante a prática de esportes. Na ginástica rítmica, esse equipamento pode ser feito de cânhamo, possuindo nós nas extremidades e sendo utilizado para a performance.
45	aparelho.n	Anatomia Geral: união de dois ou mais sistemas.
		Esportes: na ginástica, é qualquer elemento utilizado para executar os movimentos e exercícios.
46	fita.n	Vestuário: faixa estreita, de tecido natural ou sintético, us. para ornamentar ou amarrar.
		Esportes: na ginástica rítmica, é o equipamento no formato de uma banda comprida e estreita de qualquer tecido, sendo uma tira ou faixa.
46	aparelho.n	Anatomia Geral: união de dois ou mais sistemas.
		Esportes: na ginástica, é qualquer elemento utilizado para executar os movimentos e exercícios.
47	aparelho.n	Anatomia Geral: união de dois ou mais sistemas.
		Esportes: na ginástica, é qualquer elemento utilizado para executar os movimentos e exercícios.
47	fita.n	Vestuário: faixa estreita, de tecido natural ou sintético, us. para ornamentar ou amarrar.
		Esportes: na ginástica rítmica, é o equipamento no formato de uma banda comprida e estreita de qualquer tecido, sendo uma tira ou faixa.
47	maça.n	Guerra: arma formada por um cabo comprido, com uma pesada bola de ferro dentada, numa das extremidades, us. antes do advento das armas de fogo.
		Esportes: na ginástica rítmica, é o equipamento semelhante a baliza ou pino de boliche, feito de madeira ou plástico e utilizado na performance.
48	vara.n	Pesca: haste à qual se prende a linha com o anzol.
		Esportes: no salto com vara, é uma barra longa utilizada para impulsionar.
48	salto.n	Vestuário: tacão de calçado.
		Esportes: ação ou efeito de saltar; pulo.
49	atacante.n	Genérico: o que ataca; agressor.
		Esportes: jogador(a) que joga no ataque; dianteiro, avante.
49	lance.n	Pesca: ação de pescar com rede.
		Esportes: movimento, jogada.
49	lençol.n	Genérico: cada uma das duas peças de tecido, ger. leve, que se põem na cama para forrar o colchão e cobrir o corpo.
		Esportes: no futebol, é quando o jogador joga a bola por cima da cabeça do marcador e corre pelo lado em busca da bola.
50	prancha.n	Marinha: espécie de ponte entre duas embarcações ou entre uma embarcação e o cais, para passagem de pessoal.
		Esportes: equipamento utilizado no surfe em formato de uma tábua longa que flutua.

Fonte: Elaborado pelo autor (2020).

5.2 TRADUÇÃO DO CORPUS PARA A LÍNGUA-ALVO

As pesquisas geralmente buscam estabelecer padrões de referência com os quais são comparados seus testes na tentativa de se evitar avaliações subjetivas ou enviesadas. O processo de criação de um padrão de referência (*reference standard*) ou *gold standard* surge em pesquisas das áreas de Medicina e Estatística, podendo o termo ser definido como

> Qualquer avaliação clínica padronizada, método, procedimento, intervenção ou medida de validade e confiabilidade conhecidos que geralmente são considerados os melhores disponíveis, com os quais novos testes ou resultados e protocolos são comparados. (REFERENCE STANDARD, 2012)[85].

A partir da definição proposta, concebemos um padrão de referência ou *gold standard* como um método, padrão ou procedimento bastante conhecido, confiável, sendo o melhor possível, utilizado na comparação com os testes da pesquisa. Por conseguinte, faz-se necessária nesta pesquisa a criação de um *gold standard* para a avaliação dos sistemas de TM.

Antes de passarmos ao detalhamento da constituição do *gold standard* de tradução para essa pesquisa, façamos uma reflexão acerca da competência tradutória.

Acerca da competência tradutória, Hurtado Albir (2005, p. 28) postula que

> [...] é um conhecimento especializado que consiste em um sistema subjacente de conhecimentos, declarativos e, em maior proporção, operacionais, necessários para saber traduzir, que está composto de cinco subcompetências (bilíngue, extralinguística, conhecimentos sobre a tradução, instrumental e estratégica) e de componentes psicofisiológicos. (HURTADO ALBIR, 2005, p. 28).

A autora entende que a subcompetência bilíngue abarca os conhecimentos operacionais envolvidos no processo tradutório. Assim, os conhecimentos pragmático, sociolinguístico, textual, lexical e gramatical estão contemplados por essa subcompetência. A subcompetência extralinguística abrange o conhecimento de mundo, o conhecimento particular associado a aspectos culturais e enciclopédicos. Já os conhecimentos acerca da tradução integram aspectos do processo tradutório em si como os métodos utilizados, unidades de tradução, tipos de tradução e o público-alvo a quem o texto traduzido se destina. A autora apresenta a subcompetência instrumental como aquela que envolve habilidades no uso das fontes de documentação e tecnologias envolvidas no ato tradutório. Já a subcompetência estratégica se

[85] "Any standardised clinical assessment, method, procedure, intervention or measurement of known validity and reliability which is generally taken to be the best available, against which new tests or results and protocols are compared." (REFERENCE STANDARD, 2012).

coloca sobre conhecimentos operacionais que envolvem o processo tradutório. Ela se relaciona ao planejamento do ato tradutório, a detecção e solução de problemas de tradução, bem como a escolha dos melhores métodos, tudo visando à qualidade do texto traduzido. Por último, há os componentes psicofisiológicos envolvidos na competência tradutória. A autora menciona

> [...] memória, percepção, atenção e emoção; aspectos de atitude, como curiosidade intelectual, perseverança, rigor, espírito crítico, conhecimento e confiança em suas próprias capacidades, conhecimento do limite das próprias possibilidades, motivação etc.; habilidades, tais como criatividade, raciocínio lógico, análise e síntese etc. (HURTADO ALBIR, 2005, p. 29).

Com isso em mente, quando da constituição de uma tradução humana de referência, foi solicitado a um tradutor humano profissional bilíngue inglês-português, proveniente de país de língua inglesa, sendo experiente na tradução de textos específicos incluindo esportes, que realizasse a tradução do corpus de 50 sentenças da língua-fonte (português) para a língua-alvo (inglês). Essa tradução passa a ser considerada o padrão de referência e de comparação para os testes empregados nos sistemas de TM aqui desenvolvidos[86].

Características como a prática tradutória do profissional (conhecimentos sobre tradução, subcompetência instrumental e estratégica) em diversos domínios específicos incluindo os esportes (subcompetências bilíngue e extralinguística), a proficiência em português e inglês (subcompetência bilíngue) e a vivência na língua-alvo (inglês) como nativo do idioma (subcompetência extralinguística) são fatores que contribuem para que as traduções realizadas por esse tradutor humano possam ser classificadas como *gold standard*.

5.3 DESENHO EXPERIMENTAL

Nesta seção, trataremos da validação das sentenças do *gold standard* humano através de inspeção visual que verifica a preservação, na língua-alvo, dos frames evocados pelas ULs das sentenças-fonte em português. Também submetemos as sentenças em inglês apontadas como *gold standard* a nativos para averiguar a gramaticalidade e o fato de as sentenças serem possíveis na língua inglesa. Descreveremos também o funcionamento das métricas utilizadas na avaliação de TM denominadas BLEU, TER e HTER. Essas métricas serão empregadas nesta tese para avaliação dos dois sistemas aqui desenvolvidos em comparação com o sistema de TM estado da arte.

[86] Por questões de financiamento e aplicabilidade, optou-se por apenas uma tradução humana de referência (*gold standard*) com validação sintática e semântica das traduções realizadas.

5.3.1 Validação do Gold Standard Humano quanto à gramaticalidade e preservação dos Frames Evocados

Inicialmente, uma vez que o tradutor humano tenha traduzido o corpus de referência do português para o inglês, torna-se necessário submeter as sentenças traduzidas (padrão de referência) a uma verificação da preservação dos frames evocados pelas sentenças da língua-fonte nas sentenças de referência propostas na língua-alvo, além da gramaticalidade e o fato de as sentenças serem possíveis na língua. O objetivo central dessa averiguação se deve à tentativa de redução da subjetividade da tradução, dado o fato de se considerar apenas um tradutor humano de referência.

Na tentativa de se avaliarem sintaticamente as sentenças do *gold standard*, as 50 sentenças traduzidas para o inglês foram avaliadas por 4 pessoas nativas de países de língua inglesa (2 dos Estados Unidos, 1 da Inglaterra e 1 da Austrália). Foi pedido a eles que lessem as sentenças e avaliassem se estavam gramaticalmente corretas e se eram passíveis de ocorrência na língua. Alguns fatores influenciaram de forma negativa na avaliação tais como o fato de as sentenças estarem fora de um contexto maior, algumas sentenças serem de esportes específicos cujos termos um ou outro dos nativos desconheciam, além de questões particulares de variação regional do inglês, o que fez com que alguns deles sugerissem uma ou outra troca nas sentenças, mas nada que indicasse agramaticalidade ou falta de compreensão na sentença em inglês. Partindo disso, os quatro avaliadores indicaram que as sentenças estavam gramaticalmente corretas e compreensíveis na língua inglesa.

Passando a uma avaliação semântica do *gold standard*, Czulo (2017) propõe a teoria da Primazia do Frame, em que o texto original e a tradução apresentam em algum nível correspondências semânticas ao evocarem o mesmo frame ou frames próximos na rede. A partir do Apêndice B, podemos efetuar uma inspeção visual e uma quantificação acerca dos frames evocados pelas palavras de domínio específico tanto nas sentenças-fonte em português que correspondem àqueles evocados pelas ULs dos esportes nas sentenças-alvo traduzidas para o inglês (*gold standard*).

A partir dos dados e da quantificação esboçada no Apêndice B, de um total de frames específicos dos esportes evocados a partir da soma das ULs da sentença-fonte (português) e da sentença-alvo (gold standard), constatamos que tanto a sentença-fonte, quanto a sentença-alvo apresentam ULs que podem evocar o mesmo frame, havendo uma correspondência de 72,4% de frames evocados (por ULs em português e suas traduções em inglês) possíveis no domínio específico dos esportes. A porcentagem de correspondência semântica é calculada somando-se

o número de frames correspondentes evocados pelas ULs da sentença-fonte e da sentença-alvo. Faz-se uma divisão pelo número total de frames diferentes dos esportes evocados pelas ULs de ambas as sentenças. A porcentagem de 72,4% aponta que há uma elevada correspondência semântica entre as sentenças-alvo e as sentenças-fonte, constatando que escolhas tradutórias do domínio específico dos esportes podem estar semanticamente adequadas, pois as ULs traduzidas tendem a evocar os mesmos frames que as ULs das sentenças originais ou frames próximos na rede. O que faz a frequência de frames não correspondentes ser maior se deve a escolhas tradutórias de sinônimos, hiperônimos, hipônimos, além da própria ambiguidade de terminologia dentro do domínio específico, o que pode indicar que os frames dessas ULs estejam próximos na rede.

Apresentada a etapa de verificação sintático-semântica com base em frames das traduções fornecidas e avaliação de nativos em relação à gramaticalidade e aceitabilidade das sentenças, conclui-se pela adequação do *gold standard* para emprego no cálculo das métricas BLEU, TER e HTER. Passemos, portanto, a uma descrição do funcionamento da utilização de das métricas na avaliação de traduções geradas por sistemas de TM.

5.3.2 Avaliação do Desempenho dos Sistemas por Métricas

Com base na constituição de um *gold standard* de traduções para a língua-alvo, realizadas por um tradutor humano profissional nativo de país de língua inglesa e especializado em tradução de domínios específicos, incluindo os esportes, aplicou-se a verificação gramatical e semântica, exposta anteriormente, com o propósito de se constatar a preservação dos frames evocados pelas sentenças da língua-fonte na língua-alvo e a gramaticalidade/aceitabilidade das sentenças. Essa validação da preservação ou não dos frames contribui para se reduzir o caráter subjetivo concebido através de um único padrão de referência.

Havendo feito isso, passa-se à utilização de três métricas mundialmente utilizadas e reconhecidas para a avaliação dos sistemas de TM aqui propostos: BLEU, TER e HTER.

A primeira dessas métricas, a BLEU, compara n-grams da tradução candidata com n-grams da tradução de referência, independentemente da posição, insere uma penalidade por brevidade e calcula a métrica final de avaliação da tradução. Assim, sentenças candidatas a tradução que forem muito longas ou muito curtas em relação à tradução de referência recebem uma penalidade maior no cálculo final. Outro aspecto do cálculo da BLEU é que ela calcula a quantidade de *unigrams*, *bigrams*, *trigrams*, *quadrigrams* correspondentes entre a tradução analisada e o *gold standard*. Ao considerar unigrams, a métrica está avaliando mais a adequação

da tradução. Quando conta *n-grams* mais longos, o aspecto avaliado é o da fluência. Portanto, há um processo de sobreposição de *n-grams* na contabilização, fator esse que influencia no *score* final. Ao estabelecer a pontuação BLEU, é necessário que a sentença possua pelo menos um *quadrigram* correspondente para que o score seja maior que 0. Na Figura 48 podemos ver uma interpretação possível dos valores propostos pela métrica BLEU.

Figura 48 – Interpretação dos valores da métrica BLEU

Pontuação BLEU	Interpretação
< 10	Praticamente inútil
10 - 19	Difícil de compreender o sentido
20 - 29	O sentido está claro, mas há erros gramaticais graves
30 - 40	Pode ser entendido como boas traduções
40 - 50	Traduções de alta qualidade
50 - 60	Traduções de qualidade muito alta, adequadas e fluentes
> 60	Em geral, qualidade superior à humana

Fonte: *Google Cloud* - Produtos de IA e *machine learning* – AutoML - Documentação. (https://cloud.google.com/translate/automl/docs/evaluate#bleu Acesso em: 06 nov. 2020).

Nota-se a partir da Figura 48 os possíveis intervalos de valores gerados da BLEU e sua interpretação. Valores da BLEU menores que 10 indicam que a tradução está muito ruim. Entre 10 e 19, a tradução apresenta o sentido de difícil compreensão. Para *scores* entre 20 e 29, a tradução possui um sentido claro, mas erros gramaticais graves. Com a pontuação entre 30 e 40, as traduções são consideradas boas. De 40 a 50, as traduções são tidas como de alta qualidade. Com o *score* entre 50 e 60, as traduções são consideradas como possuindo uma alta qualidade, adequação e fluência. Por fim, para os valores acima de 60, a tradução é considerada superior em qualidade do que a tradução humana.

Nesta tese, a BLEU[87] foi utilizada a partir de uma aplicação online em que foi possível submeter as sentenças traduzidas pelo sistema estado da arte Google Tradutor (S-Base), o sistema enriquecido semanticamente com injeção terminológica no pré-processamento (S-Pré) e o sistema enriquecido semanticamente com injeção terminológica na pós-edição (S-Pós), comparando os resultados de tradução com uma tradução humana de referência. Vejamos na Figura 49 como são feitos os cálculos da pontuação BLEU a partir de sentenças do corpus dos esportes compilado nesta tese.

[87] https://www.letsmt.eu/Bleu.aspx Acesso em: 06 nov. 2020.

Figura 49 – Exemplo de avaliação da BLEU na tese a partir de um Avaliador Interativo online

Sentence 16	BLEU	Length ratio	Text
Human	100.00	1.00	16 - Evandro was the top scorer , having scored seven goals .
Machine	27.88	1.15	16 - The top scorer is the player Evandro , who scored seven goals .
Sentence 17	BLEU	Length ratio	Text
Human	100.00	1.00	17 - Breakaway is a technique used in road cycling .
Machine	74.19	1.00	17 - Escape is a technique used in road cycling .

Fonte: Avaliador Interativo da *BLEU Score*. (https://www.letsmt.eu/Bleu.aspx Acesso em: 06 nov. 2020).

A Figura 49 ilustra duas sentenças do corpus dos esportes compilado nesta tese (sentenças 16 e 17). A tradução de referência humana (*Evandro was the top scorer, having scored seven goals*.) representa um score BLEU de 100.0 e uma taxa de comprimento (*Length ratio*) de 1.00. A sentença traduzida gerada pelo Sistema de TM estado da arte (*The top scorer is the player Evandro, who scored seven goals*.) apresenta uma pontuação BLEU de 27.88 e uma taxa de comprimento de 1.15, por ser maior que a tradução *gold standard*. Ao observar a Figura 49, as palavras em azul compõem os n-grams que coincidem entre a tradução gerada e a de referência. Já as palavras em vermelho indicam a não correspondência entre os termos de tradução. Já no segundo exemplo, a sentença 17, cujo *gold standard* (*Breakaway is a technique used in road cycling*.) também representa um valor BLEU de 100.0 e uma taxa de comprimento de 1.0, é colocada para a avaliação de uma tradução gerada (*Escape is a technique used in road cycling*.), possuindo uma pontuação BLEU de 74.19, e taxa de comprimento de 1.0, por apresentar o mesmo tamanho da tradução *gold standard*. A partir das interpretações da métrica BLEU sugeridas na Figura 48, a tradução gerada da sentença 16, por apresentar um score de 27.88, pode ser vista como possuindo graves erros gramaticais. Mas ao analisarmos essa tradução com cuidado, ela apenas não apresenta os exatos *n-grams* da referência, mas pode ser tomada como uma boa tradução. Já no caso da tradução da sentença 17, cuja pontuação BLEU é 74.19, pode ser interpretada como possuindo qualidade superior às traduções realizadas por humanos. Entretanto, notamos que ela possui uma correspondência quase total com o *gold standard*, diferenciando-se justamente em um termo específico, que a tradução de referência apresenta como "*breakaway*", e a tradução gerada coloca como "*escape*". Ao examinarmos cautelosamente essa tradução, nota-se que "*escape*" não é um termo possível dentro do contexto dos esportes e do ciclismo, o que torna essa tradução inadequada, mesmo tendo sido avaliada com um BLEU score de 74.19. Assim, verificamos que a métrica BLEU se baseia na forma e na correspondência entre n-grams, mas se mostra inábil no que diz respeito à avaliação da

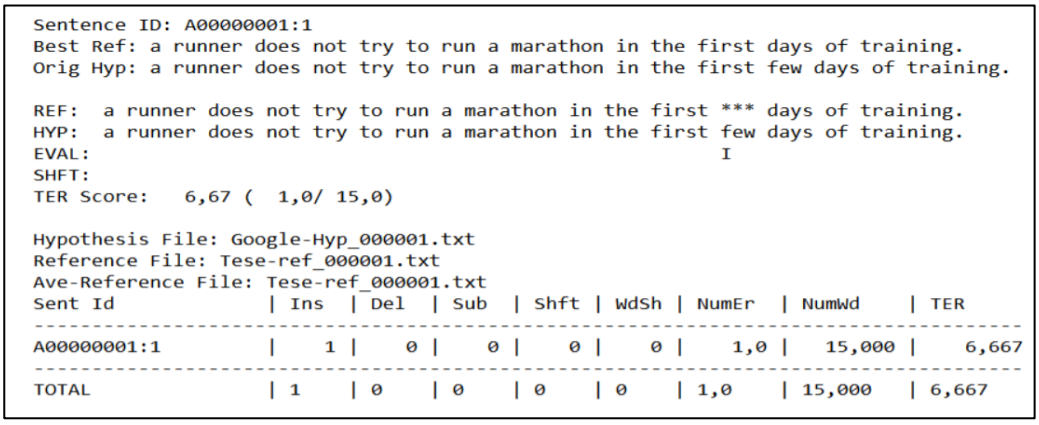
adequação semântica. A avaliação BLEU média dos sistemas de TM comparados (S-Base, S-Pré e S-Pós) pode ser observada na seção 6.2.1 e o detalhamento da pontuação BLEU por sentença do corpus de teste pode ser consultado nos Apêndices C, D e E.

As outras duas métricas debruçam-se sobre o cálculo do esforço mínimo de edição necessário para adequar formal e semanticamente a tradução gerada pelo sistema de TM a um *gold standard*. Trata-se da TER/HTER[88], as quais indicam qual é o número mínimo de edições realizadas em uma tradução para que ela corresponda, em termos de compreensão, fluência e manutenção dos significados, a um texto original de referência. Há a normalização da medida pelo comprimento médio das referências. A pontuação TER conta as edições realizadas a partir da correspondência exata entre a tradução gerada pelo sistema de TM e o *gold standard*. A versão TERp incorpora o reconhecimento de sinônimos e paráfrases nas correspondências.

As edições possuem todas o mesmo custo e incluem: a inserção, exclusão e substituição de palavras simples, além de trocas de posição em expressões ou sequências de palavras, independentemente da distância. Também são consideradas edições possíveis: a manutenção ou apagamento de caracteres de pontuação, elementos como a letra maiúscula em nomes próprios, uso ou não da marca de posse no inglês (’s), além de marcações flexionais (flexões de concordância nominal como o plural dos nomes, os adjetivos e artigos; e a concordância verbal).

A Figura 50 traz um exemplo do cálculo da TER a partir de uma sentença traduzida pelo sistema estado da arte e a comparação com o *gold standard*.

Figura 50 – Exemplo de avaliação da TER a partir do algoritmo



```
Sentence ID: A00000001:1
Best Ref: a runner does not try to run a marathon in the first days of training.
Orig Hyp: a runner does not try to run a marathon in the first few days of training.

REF:  a runner does not try to run a marathon in the first *** days of training.
HYP:  a runner does not try to run a marathon in the first few days of training.
EVAL:                                                    I
SHFT:
TER Score:   6,67 (  1,0/ 15,0)

Hypothesis File: Google-Hyp_000001.txt
Reference File: Tese-ref_000001.txt
Ave-Reference File: Tese-ref_000001.txt
Sent Id             | Ins  | Del  | Sub  | Shft | WdSh | NumEr  | NumWd    | TER
-----------------------------------------------------------------------------------
A00000001:1         |    1 |    0 |    0 |    0 |    0 |    1,0 |   15,000 |    6,667
-----------------------------------------------------------------------------------
TOTAL               | 1    | 0    | 0    | 0    | 0    | 1,0    | 15,000   | 6,667
```

Fonte: Algoritmo TER versão tercom.7.25. (http://www.cs.umd.edu/~snover/tercom/ Acesso em: 06 nov. 2020).

[88] http://www.cs.umd.edu/~snover/tercom/ Acesso em: 06 nov. 2020.

A Figura 50 traz o *gold standard* marcado como *best ref* ou *REF* (*a runner does not try to run a marathon in the first days of training*.) e a sentença traduzida gerada pelo sistema estado da arte sinalizada por *Orig Hyp* ou *HYP* (*a runner does not try to run a marathon in the first few days of training.*). Nota-se que a sentença traduzida gerada inseriu a palavra "*few*", o que indica 1 inserção. Como o tamanho do *gold standard* é de 15 palavras, o cálculo da TER se coloca como 1/15 = 6,667. Ou seja, o esforço mínimo de edições para que a tradução analisada corresponda ao *gold standard* é de 6,667. A avaliação TER média dos sistemas de TM comparados (S-Base, S-Pré e S-Pós) pode ser observada na seção 6.2.2 e o detalhamento da pontuação TER por sentença do corpus de teste pode ser consultado nos Apêndices F, G e H.

O que diferencia as métricas TER e HTER é a utilização de tradutores especialistas humanos nas edições realizadas. Na tentativa de se evitar que a tradução *gold standard* seja subjetiva, permite-se a inserção de mais de uma tradução de referência. Nesta tese, será utilizada apenas a tradução de referência de um tradutor que, após passar por verificação de correspondência semântica de frames evocados e gramaticalidade das sentenças, contribuirá para essa redução da subjetividade, dada a comparação semântica dos frames evocados pelas ULs da sentença-fonte original com as ULs da sentença-alvo traduzida. O cálculo da métrica HTER com base nas edições realizadas está contido na Equação 4.

Equação 4 – Equação do cálculo da métrica HTER

$$HTER = \frac{\text{Substituições} + \text{Inserções} + \text{Exclusões} + \text{Trocas de Posição} + \text{Alterações}}{\text{Palavras de Referência}}$$

Fonte: Elaborado pelo autor (2020) a partir de https://languagelog.ldc.upenn.edu/nll/?p=193 Acesso em: 13 out. 2020.

A opção pelo uso da HTER como métrica de avaliação de TM nesta tese ocorre porque outras métricas não dão conta de captar os aspectos semânticos da tradução ou são inviáveis por condições de aplicabilidade. A métrica BLEU não se mostra eficaz por focar unicamente em aspectos posicionais e sintáticos, desconsiderando as peculiaridades de significado em textos maiores. A MEANT, métrica que opera com questões semânticas, não se coloca como viável pois depende da anotação de um *PropBank* para português, dando importância aos papéis semânticos e à estrutura de evento da sentença-fonte original e da sentença-alvo traduzida. A TER foca mais na forma e deixa de lado alguns aspectos semânticos da tradução. Por exemplo, ela contaria como edições as trocas necessárias para adequar uma dada construção linguística

que mantém a semântica da tradução por outra equivalente escolhida no *gold standard*. Dado isso, a HTER é considerada a métrica de avaliação de TM que melhor atende os critérios específicos para avaliação de TM de domínio específico tanto nos aspectos sintáticos quanto semânticos. Apesar de possuir uma grande limitação no fato de ser dispendiosa e necessitar de humanos no processo, e apresentar ruído, ou subjetividade na tradução, a verificação prévia da preservação dos frames do original com os da tradução e o fato de conseguirmos os tradutores/revisores humanos, tornam a aplicação da HTER como um método de avaliação de TM exequível neste trabalho.

Analisemos o funcionamento e o cálculo da HTER apoiados nas sentenças e traduções contidas nos exemplo (6) a (11).

(6) Fuga é uma técnica utilizada no ciclismo de estrada. (Sentença-fonte original)
(7) Breakaway is a technique used in road cycling. (*Gold Standard*).
(8) Escape is a technique used in cycling of road. (Linguatec Tradutor[89]).
(9) Breakaway is a technique used in cycling of road. (Tradutor/Editor Humano 2: Linguatec editada) [1 substituição, 1 apagamento, 1 troca de posição].
(10) Breakaway is technique used in road cycling. (S-Pós[90]).
(11) Breakaway is technique used in road cycling. (Tradutor/Editor Humano 3: S-Pós) [0 mudanças].

Apoiado nos exemplos acima, temos a sentença original em português em (6). Em (7), temos hipoteticamente a sentença que será o padrão de referência de tradução, realizada por um tradutor humano nativo especializado, com verificação quanto à similaridade semântica com a sentença original através da métrica F-SEM. Em (8) e (10), temos as traduções oferecidas para a sentença original por dois sistemas diferentes, sendo (8) por um tradutor online Linguatec, e (10) pelo S-Pós proposto por esta tese. Em (9) temos o resultado do número mínimo de edições que um revisor X realizou em (8) para que essa tradução fosse compreensível e fiel semanticamente ao sentido proposto em (7). Em (11), notamos as edições necessárias realizadas minimamente por um editor Y para a tradução de (10) fosse fluente e compatível semanticamente com o que foi proposto em (7).

[89] https://www.linguatec.de/personal-translator-demo/
[90] O Sistema de TM, enriquecido semanticamente com frames e qualia, com injeção terminológica na pós-edição é apresentado e descrito na seção 6.1.3.

Colocando os dados de (9) e (11) na Equação 5 e na Equação 6, respectivamente, analisemos:

Equação 5 – HTER aplicada sobre o resultado da LINGUATEC de (9)

$$HTER = \frac{1\ (\text{substituição}) + 1\ (\text{apagamento}) + 1\ \text{troca de posição}}{8\ (palavras\ da\ tradução\ de\ referência)} = \frac{3}{8} = 0{,}375$$

Fonte: Elaborado pelo autor (2020)

Equação 6 – HTER aplicada sobre o resultado do S-Pós de (11)

$$\text{HTER} = \frac{0\ (\text{modificações})}{8\ (\text{palavras da } tradução\ de\ referência)} = \frac{0}{8} = 0$$

Fonte: Elaborado pelo autor (2020)

A métrica HTER postula que quanto menor for o número de edições possíveis, maior será o grau de similaridade entre a sentença original na língua-fonte e sua tradução na língua-alvo. A partir do que observamos na Equação 5 e na Equação 6, notamos que a HTER é de 0,375 para a sentença traduzida pela Linguatec, e de 0 para a sentença traduzida pelo S-Pós proposto nesta tese. Portanto, para essa comparação específica, notamos que houve uma boa correspondência de tradução entre o que foi gerado pelo S-Pós e o *gold standard*, deixando a tradução gerada pela Linguatec como não tão adequada.

Com a implementação da HTER, ocorre a submissão das sentenças traduzidas pelos S-Base (simples com uso de RNNs), S-Pré (melhorado semanticamente com injeção terminológica no pré-processamento) e S-Pós (enriquecido semanticamente com injeção terminológica na pós-edição), para a avaliação de tradução e desempenho dos sistemas. No âmbito desta tese, como descrito acima, as sentenças traduzidas por cada um dos três sistemas são submetidas a três tradutores humanos especializados diferentes, que desempenham um papel de revisores/editores. Assim, eles devem propor versões finais para o seu grupo de traduções, realizando o mínimo de edições possíveis para que suas sentenças traduzidas se aproximem semanticamente das sentenças traduzidas pelo *gold standard*, ou seja, através de traduções compreensíveis, aceitáveis no domínio específico dos esportes e que mantenham o mesmo sentido das traduções do padrão de referência. O cálculo da quantidade e dos tipos de edição foi feito manualmente de forma independente por dois pesquisadores diferentes que

falam a língua inglesa, sendo revisado por um terceiro analista. Posto isso, a HTER é calculada e estabelece-se uma medida ou métrica geral de avaliação dos sistemas de tradução, fundamentada pela comparação, realizada por humanos entre os resultados gerados pelos algoritmos de TM e a tradução de referência. A avaliação HTER média dos sistemas de TM comparados (S-Base, S-Pré e S-Pós) pode ser observada na seção 6.2.3 e o detalhamento da pontuação HTER por sentença do corpus de teste pode ser consultado nos Apêndices I, J e K.

Passemos ao Capítulo 6 com a apresentação do sistema de desambiguação (DAISY), dos modelos de TM aqui propostos, sua avaliação e a discussão dos resultados.

6. MODELOS DE TRADUÇÃO POR MÁQUINA HÍBRIDOS BASEADOS EM FRAMES

Este capítulo introduz o Sistema de Desambiguação (DAISY) e ilustra a proposição de dois modelos de TM enriquecidos semanticamente com frames e relações qualia. Além disso, são mostrados os testes e avaliações dos modelos de TM, sendo feita uma discussão detalhada acerca do seu desempenho.

6.1 PROPOSIÇÃO DOS MODELOS DE TM

Nesta seção, apresentaremos dois modelos híbridos de TM semanticamente enriquecidos com a Semântica de Frames e as Relações Qualia desenvolvidos no âmbito desta tese, sejam eles: (i) um Sistema de Tradução por Máquina com Injeção Terminológica no Pré-processamento das sentenças a serem traduzidas; e (ii) um Sistema de Tradução por Máquina com Injeção Terminológica na etapa de Pós-edição. Abordemos primeiramente, porém, um Sistema de Desambiguação de Frames (DAISY) utilizado como etapa nos sistemas de TM aqui desenvolvidos.

6.1.1 Sistema de Desambiguação (DAISY)

Os modelos de TM propostos nesta tese fazem uso da base de dados da FrameNet Brasil. Concebemos uma base de dados como uma representação semântica do conhecimento, neste trabalho, voltada ao domínio específico dos esportes. As sentenças submetidas aos sistemas de TM podem conter palavras que constem dessa base. Para que os sistemas de TM funcionem adequadamente, o primeiro passo é a realização de uma associação entre as palavras da sentença na língua-fonte com palavras (ou representações semânticas) modeladas na base de dados. A partir de relações de equivalência de tradução existentes entre palavras da língua-fonte e seus equivalentes na língua-alvo na base de dados, o sistema é capaz de realizar associações entre as palavras da sentença a ser traduzida com sua representação semântica na base e propor seus equivalentes de tradução adequados no domínio específico dos esportes. Entretanto, para que tal proposição possa ocorrer, primeiramente é necessário que o sistema seja capaz de identificar os frames aos quais os itens lexicais estão associados e, para o caso de lemas polissêmicos, desambiguar-lhes o sentido. Portanto, para que essa associação seja feita de forma correta,

utiliza-se como etapa dos sistemas de TM aqui propostos um sistema de desambiguação denominado DAISY.

O Sistema de Desambiguação DAISY (*Disambiguation Algorithm for Inferring the Semantics of Y*)[91] foi desenvolvido pela FrameNet Brasil como um algoritmo que elenca todos os frames ativados na base de dados da FN-Br por uma certa palavra da sentença. Ele utiliza técnicas de desambiguação (WSD – *Word Sense Disambiguation*) adaptadas ao uso com frames. O processo desempenhado pelo DAISY aplica um modelo de atribuição de rótulos chamado Campos Aleatórios Condicionais (CRF – *Conditional Random Fields*). Sutton e Mccallum (2012) apresentam os CRFs como

> [...] uma forma de combinar as vantagens da classificação discriminativa e modelagem gráfica, combinando a capacidade de modelar compactamente saídas multivariadas **y** com a capacidade de impulsionar um grande número de recursos de entrada **x** para a previsão. (SUTTON; MCCALLUM, 2012, p. 269)[92].

Assim, a partir da definição proposta por Sutton e McCallum (2012), compreendemos os CRFs como um método de modelagem estatística utilizado para o reconhecimento de padrões, a aprendizagem de máquina e a previsão estruturada (*structured prediction*). Sutton e McCallum (2012) ainda propõem que os métodos de previsão estruturada

> [...] são essencialmente uma combinação de classificação e modelagem gráfica, combinando a capacidade de modelar dados multivariados de forma compacta com a capacidade de realizar uma previsão usando grandes conjuntos de recursos de entrada. (SUTTON; MCCALLUM, 2012, p. 353)[93].

A respeito de processos da atribuição de classes de palavras, lematização e análise de dependências entre as palavras, o DAISY utiliza o UDPipe[94], sendo este um pipeline treinável, independente de línguas, aplicado em processos de tokenização, atribuição de classes de palavra, lematização e análise de dependência de arquivos anotados no formato CoNLL-U.

O sistema do DAISY se baseia nas relações existentes na base da FrameNet Brasil, tais como aquelas entre os frames evocados pelas palavras que ocorrem na sentença. A partir dessas relações, há a atribuição de um peso de ativação para os nós na rede, tais como os frames, por

---

[91] https://github.com/FrameNetBrasil/daisy

[92] "[…] a way of combining the advantages of discriminative classification and graphical modeling, combining the ability to compactly model multivariate outputs y with the ability to leverage a large number of input features x for prediction." (SUTTON; MCCALLUM, 2012, p. 269).

[93] "Structured prediction methods are essentially a combination of classification and graphical modeling, combining the ability to compactly model multivariate data with the ability to perform prediction using large sets of input features." (SUTTON; MCCALLUM, 2012, p. 353).

[94] http://ufal.mff.cuni.cz/udpipe

exemplo. Quanto maior o caminho percorrido entre frames, menor o índice de ativação. Havendo uma palavra polissêmica, o sistema é capaz de desambiguá-la ao propor qual frame ela evoca com maior ativação a partir do contexto de ocorrência na sentença. Vejamos melhor sobre o funcionamento desse sistema de pesos e ativação denominado Ativação Propagada (*Spreading Activation*).

O modelo do DAISY é implementado como um algoritmo de ativação propagada (SA - *spread activation*) que consiste em técnicas de processamento aplicadas em uma estrutura de dados em rede. Crestani (1997) elucida que a ativação propagada ocorre em uma rede formada por nós interligados através de arcos (links ou relações). Há iterações na ativação dos nós e, uma vez que um nó é ativado, outros nós anteriores a ele também foram ativados. São atribuídos nomes aos nós que reproduzem objetos ou suas características. Já os arcos simbolizam relações entre os nós, podendo eles receberem rótulos como nomenclatura ou até mesmo serem ponderados através de pesos. Os arcos normalmente possuem uma direção específica, fato este que influencia no rótulo ou peso atribuído a eles. Analisemos como se realiza o cálculo da ativação propagada através da Equação 7.

Equação 7 - Cálculo da ativação propagada

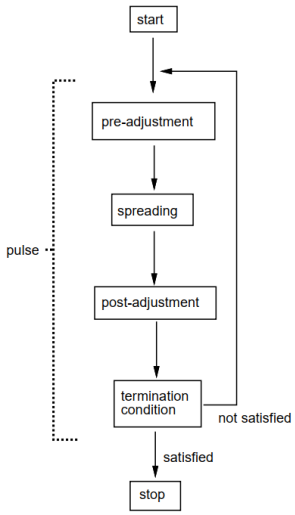
$$O_j(p) = f(A_j(p))$$

Fonte: Identificação Automática de Construções de Estrutura Argumental: um experimento a partir da modelagem linguístico-computacional das construções transitiva direta ativa, ergativa e de argumento cindido. (ALMEIDA, 2016, p. 52).

A Equação 7 ilustra o funcionamento de um algoritmo de ativação propagada. A ativação dos nós se espalha para os nós adjacentes, em função do valor atual da ativação e dos níveis de ativação estabelecidos entre os arcos que conectam os nós próximos. Tomando cada iteração **p**, um nó **j** possui uma ativação representada por **Aj(p)** e gera uma saída **Oj(p)**, sendo esta uma função em relação ao seu nível de ativação. Os valores de entrada e dos pesos geralmente são números reais, podendo ser binários (0 ou 1).

Crestani (1997) ilustra as etapas de um sistema de ativação propagada conforme vemos na Figura 51.

Figura 51 – Modelo de Ativação Propagada (SA - *Spreading Activation*)



Fonte: Application of Spreading Activation Techniques in Information Retrieval. (CRESTANI, 1997, p. 9).

A Figura 51 traz o funcionamento básico de um algoritmo de SA, o qual consiste em uma fase inicial (*start*) em que pulsos (*pulses*) são lançados na rede, ativam os nós e vão se espalhando (*spreading*) através dos arcos (links ou relações) até que condições de parada (*termination condition*) satisfaçam as restrições ou características programadas inicialmente no sistema. Uma vez que essas condições de parada são atingidas (*satisfied*), o sistema interrompe a propagação e para (*stop*). Caso as condições não sejam atendidas (*not satisfied*), o pulso avança ou retorna pela rede. Sistemas mais complexos podem incluir estágios de pré-ajuste (*pre-adjustment*) ou pós-ajuste (*post-adjustment*). Nessas fases específicas, condições podem ser inseridas no sistema, fazendo com que os níveis de ativação caiam, o que contribui para que a ativação não fique retida em um certo ponto. Há uma associação dessa queda dos níveis de ativação com uma "perda do interesse do sistema" conforme as condições ou restrições que estão modeladas na rede.

Durante o funcionamento do DAISY, há a formação de um agrupamento, ou *cluster*, com o conjunto de frames evocados e considerando a classe e a sintaxe das palavras na sentença via UDPipe. Em um estágio inicial, a similaridade entre os frames é avaliada através do *cluster*, ou seja, se um mesmo frame está sendo evocado por mais de uma UL ou não. Em uma dada sentença, uma palavra se associa a uma UL da base e ativa um frame (nó). As características da relação semântica estabelecida (arco), seja ela frame a frame, EF-frame ou qualia, representam um contexto de aplicação e possuem pesos ou valores associados à relação, os quais foram estimados a partir daquilo que tais relações representam em termos cognitivos. No caso dos

frames diretamente evocados, o peso associado é de 1,0. Se os frames foram ativados via relação de Herança entre frames, o nível de ativação será 1,0. Para relação de Perspectiva entre frames, o seu peso será de 0,9. Já os frames ativados que estiverem em relações de Subframe ou Superframe possuem um peso de 0,7. Por último, os frames associados via relação de Uso apresentam um peso de 0,6. Se os frames pertencem ao domínio m.knob (frames de esportes e turismo), eles ganham um peso de 5. Outro tipo de relação semântica que também contribui para processos de desambiguação e para aumento nos níveis de ativação é a relação EF a frame. Assim, esse tipo de relação possui um peso de 0,5.

Refletindo acerca das relações qualia modeladas entre as ULs ativadas a partir das palavras da sentença, quanto mais palavras da sentença a ser traduzida estiverem associadas a ULs da base que possuem relações qualia entre elas ou de forma indireta, maiores serão os níveis de ativação para os frames que essas ULs evocam. Assim, as relações qualia podem também ser associadas aos arcos da ativação propagada e as ULs ligadas por elas podem ser vistas como nós. O peso atribuído para cada relação qualia existente entre as ULs é de 0,9.

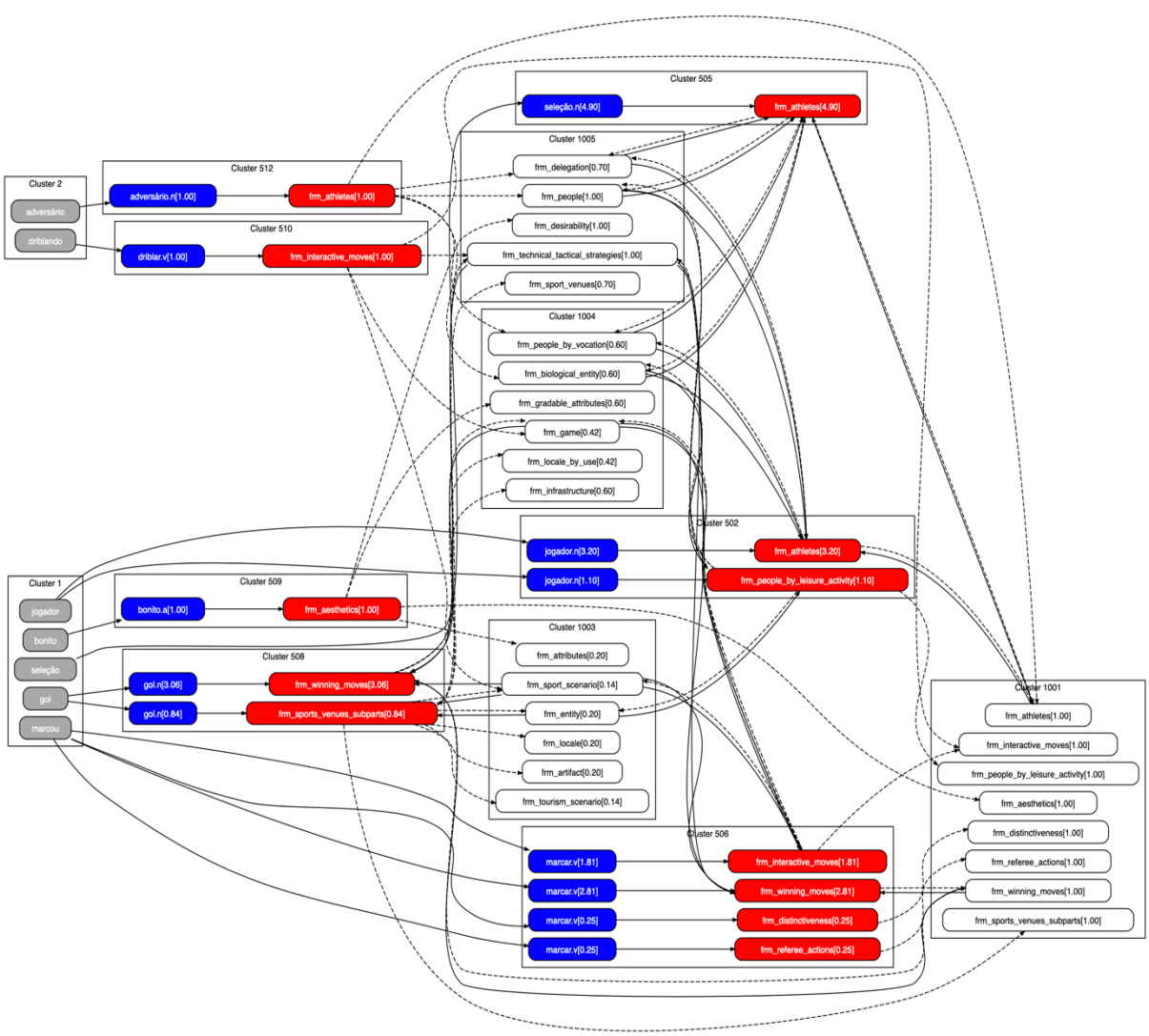
O Sistema DAISY também oferece um tratamento diferenciado para Expressões Multipalavras (*Multiword Expressions* - MWE). Caso as palavras da sentença componham MWEs existentes na base de dados da FN-Br como ULs, é atribuído um peso de 10 para os frames em que essas MWEs estão modeladas, contribuindo para que o sistema reconheça nas sentenças expressões multipalavras e não considerando apenas as palavras da expressão soltas.

Passando à análise do DAISY, tomemos a sentença-exemplo específica dos esportes em (12).

(12) O jogador da seleção marcou um gol bonito driblando o adversário.

A partir da sentença em (12), vejamos a representação gráfica e o funcionamento do Sistema de Desambiguação DAISY na Figura 52.

Figura 52 – Funcionamento do Sistema de Desambiguação DAISY



Fonte: DAISY na FrameNet Brasil. (http://server2.framenetbr.ufjf.br:8010/index.php/daisy/main#).

A Figura 52 demonstra o funcionamento do sistema de desambiguação DAISY. Inicialmente, o algoritmo extrai as palavras da sentença (12) que evocam frames na base de dados da FN-Br e as separa em dois *clusters*. Nesses dois clusters iniciais, as palavras são marcadas com o fundo na cor cinza. São formados dois *clusters* a partir da análise de dependência gerada pelo UDPipe. No *cluster* 1, notamos a extração das palavras jogador, bonito, seleção, gol e marcou. Já no cluster 2, temos o agrupamento das palavras adversário e driblando. Em azul, estão as ULs com as quais essas palavras se associam na base de dados, sendo elas **jogador.n**, **bonito.a**, **seleção.n**, **gol.n**, **marcar.v**, **adversário.n** e **driblar.v**. A detecção da classe de palavra e busca pela UL correspondente àquela classe de palavra na sentença ocorre também devido à implementação do UDPipe, ferramenta que realiza a

tokenização, capta as relações de dependência entre os sintagmas e rotula as palavras da sentença com a classe de palavras com base na ocorrência sintática das palavras. Em vermelho, estão os frames evocados por cada UL, ou seja, "frames diretos". Por último, com o fundo branco, temos os frames que estão relacionados na rede da FN com os "frame diretos", sendo portanto considerados "frames indiretos". Seguindo a técnica de SA, quanto mais distantes estão os frames na rede e quanto mais afastadas ou indiretas relações qualia que ligam as ULs, menores serão os níveis de ativação ou pesos dos frames.

No exemplo da Figura 52, a UL **adversário.n** ativa o frame Atletas (*Athletes*) com um peso 1.00. A UL **seleção.n** também ativa o frame Atletas mas com um peso 4.90. Por último, a palavra jogador se associa a duas ULs **jogador.n**, evocando os frames Atletas (*Athletes*) e Pessoas_por_atividade_de_lazer (*People_by_leisure_activity*). Entretanto, como a UL **jogador.n** do frame Atletas está no domínio específico dos esportes, possui relações qualia diretas (constitutivo: membro de **seleção.n**, constitutivo: relaciona-se com **adversário.n**, télico: tem como atividades **driblar.v** e **marcar.v**) e indiretas (formal: é um tipo de **atleta.n**, constitutivo: **adversário.n** relaciona-se com **atleta.n**) com outras ULs da sentença, o frame Atletas é ativado com um peso de 3.20, ficando o frame Pessoas_por_atividade_de_lazer com uma ativação de 1.10. A UL **driblar.v** evoca o frame Jogadas_interativas (*Interative_moves*) com um peso de 1.00 e **bonito.a** evoca o frame Estética (*Aesthetics*) com um nível de ativação de 1.00. Por outro lado, a palavra gol da sentença evoca dois frames, Jogadas_pontuadas (*Winning_moves*) com um nível de ativação 3.06, e Subpartes_de_instalações_esportivas (*Sports_venues_subparts*) com um peso de 0.84). A desambiguação de ULs para a palavra gol e maior ativação empregada adequadamente para o frame de Jogadas_pontuadas ocorre devido às relações entre frames na base, além das relações qualia diretas (télico: **gol.n** é uma atividade do **jogador.n**) e indiretas (télico: **jogador.n** tem como atividade **marcar.v**) existentes entre as ULs evocadas pelas palavras da sentença. Por último, temos a palavra marcar que se associa a diferentes ULs **marcar.v**, evocando os frames Jogadas_interativas (*Interactive_moves* – ativação:1.81), Jogadas_pontuadas (*Winning_moves* – Ativação: 2.81), Distinção (*Distinctiveness* – ativação: 0.25), e Ações_do_árbitro (*Referee_actions* – nível de ativação: 0.25). O processo de desambiguação entre as possíveis ULs **marcar.v** ocorre da mesma forma do que descrevemos anteriormente para as outras ULs e, havendo relações entre frames e qualia entre as ULs associadas pelas palavras contidas na sentença, o processo

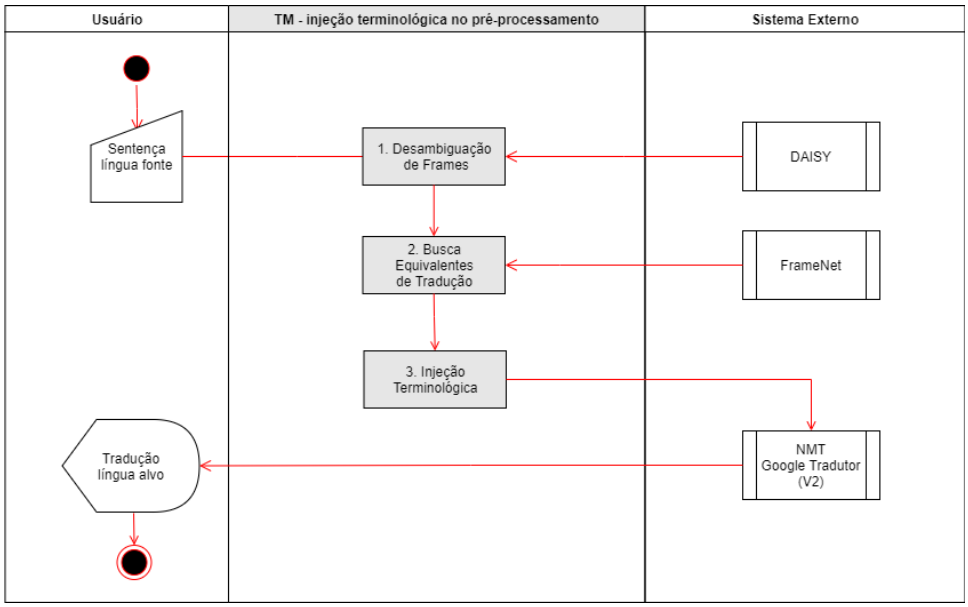
de desambiguação atribui um peso de ativação de 2.81 para o frame `Jogadas_pontuadas`, fazendo com que o sentido mais adequado para marcar no contexto da sentença seja o definido pela UL nesse frame.

Concluindo, os pesos para os frames indiretos relacionados na base "voltam" para os frames diretos, aumentando sua energia de ativação. Por fim, o frame direto com maior nível de ativação será aquele oferecido pelo DAISY como o correto para a palavra de uma dada sentença em contexto, havendo a associação da palavra da sentença com a UL que evocou esse frame. Passemos então para a subseção 6.1.2, em que se apresenta o sistema de TM com injeção terminológica realizada na etapa de pré-processamento.

6.1.2 Sistema de Tradução por Máquina com Injeção Terminológica no Pré-processamento

Um dos sistemas desenvolvidos nesta pesquisa utiliza a injeção terminológica na etapa de pré-processamento e funciona conforme o *pipeline* ou diagrama de execução ilustrado abaixo na Figura 53**Erro! Fonte de referência não encontrada.**.

Figura 53 – Diagrama das etapas do sistema de tradução com injeção terminológica no pré-processamento



Fonte: Elaborado pelo autor (2020).

Conforme se observa na Figura 53, o sistema de TM com Injeção terminológica no pré-processamento possui quatro etapas de funcionamento, contando com a utilização de três

sistemas externos. Na primeira etapa, a sentença de origem, em português, por exemplo, é inserida no sistema de Desambiguação DAISY. A partir das relações entre frames, EF-frame e das relações qualia modeladas entre palavras da sentença, o sistema realiza a desambiguação de frames, selecionando o frame que possuir maior nível de ativação e implementando um refinamento semântico dos dados. Na segunda etapa, o sistema faz uma busca de equivalentes de tradução na base de dados da FrameNet Brasil a partir das relações de equivalência de tradução estabelecidas entre as ULs do domínio específico dos Esportes. Assim, a partir da associação estabelecida entre uma palavra da sentença a uma UL da base e, consequentemente, a evocação de um frame *F*, faz-se uma busca pelo seu equivalente em inglês no mesmo frame *F*, e há uma seleção dos possíveis equivalentes para aquela palavra. Na terceira etapa, há a Injeção Terminológica de um dos termos possíveis em inglês para uma dada palavra em português em uma sentença híbrida intermediária no processo tradutório denominada injetada (*injected*). Na quarta e última etapa, a sentença intermediária injetada é submetida ao Google Tradutor (Versão NMT – V2 – com *Transformers* Universais), que realiza a tradução para a língua-alvo, o inglês por exemplo, e gera uma tradução semanticamente enriquecida, com melhorias lexicais de domínio específico a partir de frames e relações qualia.

A Figura 54 traz o funcionamento do Sistema de TM com Injeção Terminológica no Pré-processamento a partir de uma sentença-exemplo do domínio dos esportes. A sentença em questão a ser traduzida do português para o inglês é dada em (13) e o funcionamento do processo tradutório é fornecido na Figura 54.

(13) No rugby, o abertura é o jogador mais habilidoso do time.

A Figura 54 pode ser dividida em quatro partes. Na primeira parte, notamos as palavras e respectivas ULs extraídas da sentença-fonte e os frames que evocam com seus pesos (*weights*), ou seja, os níveis de ativação. Com o intuito de ilustrar esses pesos ou níveis de ativação, temos a palavra abertura que, na base de dados da FN-Br, pode se associar a três ULs e, consequentemente, evocar três frames: `Jogadas_pontuadas` (*`Winning_moves`*) com um nível de ativação de 6.23, `Cerimônias` (*`Ceremonies`*) com um peso de 5.33, e `Atletas_por_posição` (*`Athletes_by_position`*), com a ativação de 13.83. A diferença no nível de ativação dos frames a partir da palavra polissêmica abertura e suas ULs associadas se deve ao fato de, na sentença de origem, haver outras palavras e ULs associadas ligadas com a UL **abertura.n** (`Atletas_por_posição` – *`Athletes_by_position`*),

o que proporciona um aumento do nível de ativação. A segunda parte da Figura 54, marcada pela palavra vencedor (*winner*), separa as ULs cujos frames tiveram um maior índice de ativação e, a partir das relações de equivalência de tradução estabelecidas entre ULs na base de dados, o sistema oferece a UL equivalente em língua inglesa. Na camada vencedor (*winner*), notamos que as ULs do domínio específico dos esportes que apresentaram frames com um maior nível de ativação foram: **rugby.n** (frame: `Esportes – Sports`, ativação: 7.80, equivalente: ***rugby*.n**), **abertura.n** (frame: `Atletas_por_posição – Athletes_by_position`, ativação 13.83, equivalente: ***fly-half*.n**), **jogador.n** (frame: `Atletas – Athletes`, ativação 13.62, equivalente: ***player*.n**) e **time.n** (frame: `Atletas – Athletes`, ativação 13.02, equivalente: ***club*.n**). Tendo o sistema feito essa busca de equivalentes de tradução, podemos ver, na terceira parte da Figura 54, marcada pela palavra qualias, as relações qualia modeladas entre as possíveis ULs associadas a partir das palavras da sentença. Nota-se que há relações qualia entre as ULs **rugby.n** (`Esportes`), **abertura.n** (`Atletas_por_posição`), **jogador.n** (`Atletas`) e **time.n** (`Atletas`), fato este que contribuiu para que o frame com maior ativação para a UL **abertura.n** fosse `Atletas_por_posição` (peso: 13.83) e não `Jogadas_pontuadas` (peso: 6.23) ou `Cerimônias` (peso: 5.33).

Os equivalentes de tradução são separados e parte-se para a última etapa ilustrada na Figura 54 que é a fase intermediária de injeção terminológica em que os equivalentes de tradução são injetados na sentença de origem antes de serem submetidas ao Google Tradutor (Versão NMT – V2 – com Transformers Universais). A sentença intermediária também pode ser denominada injetada (*injected*) e nela já houve uma tradução prévia dos termos específicos dos esportes a partir da desambiguação realizada anteriormente. Logo, a sentença original em português (13) possui uma versão híbrida injetada em (14) e sua tradução final adequada ao domínio dos esportes e semanticamente melhorada com frames e relações qualia em (15).

(14) en o rugby, o fly-half é o player mais habilidoso de o club.

(15) In rugby, the fly-half is the club's most skilled player.

Conforme vai se demonstrar na parte de avaliação, o sistema de pré-processamento, por fazer a substituição e submeter uma sentença híbrida injetada, parte em português e parte em inglês, faz com que o desempenho do tradutor piore no que concerne às flexões nominais e verbais (plural, concordância), algumas trocas de preposição e apagamento de artigos.

Entretanto, na seção de avaliação, verificar-se-á a melhoria lexical e adequação semântica ao domínio específico dos esportes proporcionada por esse sistema de TM com injeção terminológica em pré-processamento. A partir de um diagnóstico prévio das perdas causadas nas propriedades flexionais da tradução, alguns problemas na tradução de preposições e artigos (vide Apêndice A), há uma motivação para o desenvolvimento de outro sistema de TM que busca propor sentenças traduzidas enriquecidas semanticamente com frames e relações qualia que não apresentem as perdas tradutórias descritas.

Figura 54 – Exemplo de funcionamento do sistema de tradução com injeção terminológica na etapa de pré-processamento

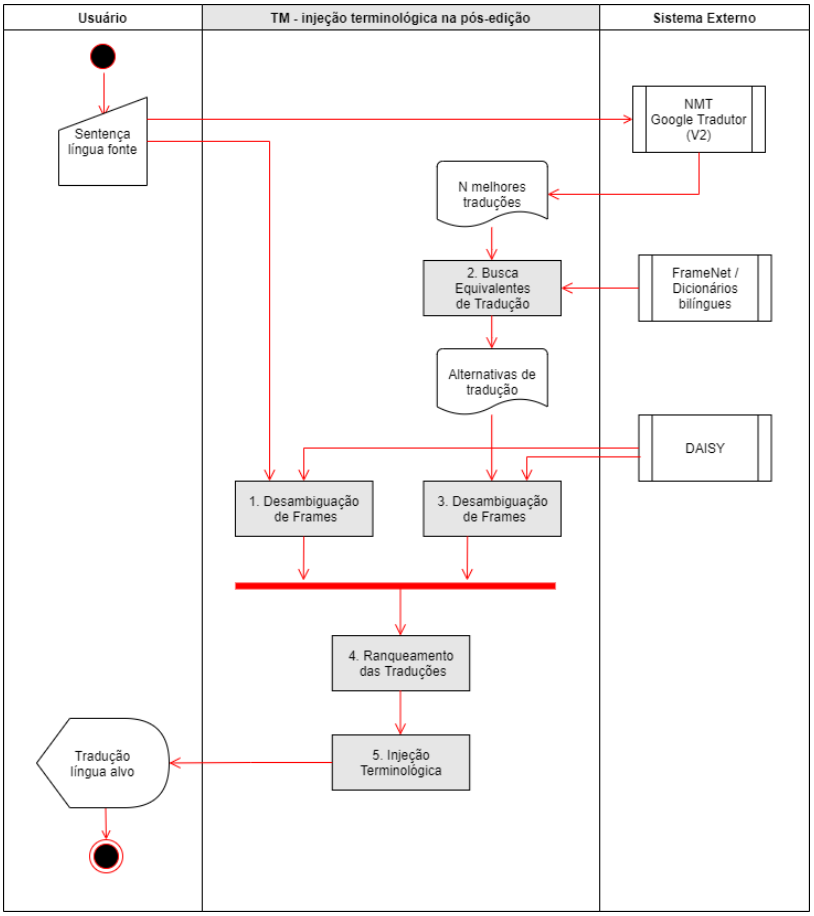
Fonte: Sistema de TM com injeção em pré-processamento da FN-Br. (http://server2.framenetbr.ufjf.br:8010/index.php/daisy/main/injection#).

Passemos então à subseção 6.1.3 com a descrição de um outro sistema de TM que realiza a injeção terminológica na etapa de pós-edição.

6.1.3 Sistema de Tradução por Máquina com Injeção Terminológica na Pós-edição

Outro sistema de TM desenvolvido no âmbito desta tese e ora apresentado é um sistema de TM enriquecido semanticamente com frames e relações qualia com o processo de injeção terminológica realizada na etapa de pós-edição. A Figura 55**Erro! Fonte de referência não encontrada.** traz um gráfico que ilustra as etapas de funcionamento desse sistema.

Figura 55 – Diagrama das etapas do sistema de tradução com injeção terminológica na pós-edição



Fonte: Elaborado pelo autor (2020).

Observando o pipeline ou esquema ilustrado na Figura 55**Erro! Fonte de referência n ão encontrada.** detalhemos as etapas de funcionamento desse sistema. Inicialmente, na primeira etapa, uma sentença na língua-fonte é submetida ao sistema de desambiguação de

frames (DAISY). Como já descrito anteriormente, o sistema de desambiguação utiliza o *parser* UDPipe (com as categorias linguísticas utilizadas pelas *Universal Dependencies*) com a intenção de se realizar um tratamento linguístico e de categorização prévia de classes de palavras (*Part of Speech* – POS) para as ULs das sentenças a serem traduzidas e então rodar o sistema de desambiguação de frames. A partir das relações frame a frame, EF a frame e relações qualia entre as ULs associadas a partir de palavras da sentença, o sistema ativa os frames evocados na rede da FrameNet Brasil com um determinado nível de ativação. Caso a palavra não seja polissêmica, há uma atribuição de peso de ativação para um dado frame. Entretanto, tratando de ULs ambíguas, mais de um frame é ativado na rede. Partindo disso, quando mais relações entre frames houver e, quanto mais relações qualia entre ULs da sentença existirem, maior será o nível de ativação para um dado frame, ocorrendo a desambiguação de frames e selecionando o frame mais adequado para a UL evocada pela palavra da sentença contextualmente. As informações geradas pelo DAISY são armazenadas para uso posterior.

Na segunda etapa, a sentença da língua-fonte é submetida ao sistema NMT do Google Tradutor V2 que utiliza *Transformers* Universais, previamente treinado para um dado par de sentenças, sendo geradas as N melhores (neste modelo, as 5 melhores) traduções possíveis na língua-alvo para a sentença-fonte.

Na terceira etapa, as melhores traduções passam por um processo de ranqueamento. Há a utilização de um dicionário bilíngue (português-inglês) após o sistema NMT do Google Tradutor V2 gerar as traduções. A aplicação do dicionário tem o propósito de contribuir para o reconhecimento de que palavra foi traduzida da língua-fonte por qual palavra na língua-alvo. As traduções geradas geram um conjunto de frames desambiguados com o DAISY para cada tradução. Assim, a partir dessa associação, o sistema consegue detectar problemas de tradução e injetar corretamente as traduções nos lugares adequados da sentença traduzida. Com os conjuntos gerados, cada tradução tem seus frames pareados com o conjunto de frames desambiguados da sentença de origem (armazenados anteriormente). O número de frames iguais entre os conjuntos da tradução e da sentença de origem é o ranking utilizado para a avaliação semântica. Dentre as traduções, aquela que obtiver o melhor ranking, continuará para o último passo.

Na quarta e última etapa, a tradução selecionada pela etapa 3 passa pelo processo de injeção terminológica realizada com o um algoritmo de injeção terminológica (FNTi), em que é feita a tentativa de substituição das ULs pelas suas equivalentes de tradução na língua-alvo, gerando assim, sentenças traduzidas enriquecidas semanticamente com frames e relações qualia e com equivalentes de tradução mais adequados para um domínio específico, seja ele o dos

Esportes. Dessa forma, esse sistema de TM desenvolvido nesta tese e enriquecido semanticamente e com injeção terminológica na pós-edição apresenta-se como inovador e como uma alternativa aos sistemas de NMT convencionais que utilizam RNNs para o tratamento de TM para sentenças de domínios específicos. Veremos e discutiremos os resultados e a avaliação desse sistema de TM na seção de a 6.2.

6.2 AVALIAÇÃO DOS MODELOS DE TM

Esta seção apresenta os resultados gerais apresentados por ambos os sistemas de TM (com injeção terminológica no pré-processamento e na pós-edição) em comparação com o sistema de TM estado da arte (Google Tradutor), além de estruturar uma sistematização de avaliação dos sistemas com as métricas BLEU, TER e HTER, e discutir detalhadamente os resultados gerados, suas melhorias e problemas encontrados.

As 50 sentenças do corpus de teste específico dos esportes foram traduzidas pelo sistema de TM estado da arte (S-Base) e por dois sistemas de TM semanticamente enriquecidos com frames e relações qualia, sendo o primeiro com injeção terminológica no pré-processamento (S-Pré), e o segundo com a injeção terminológica na pós-edição (S-Pós). Foi utilizada uma tradução humana feita por um tradutor bilíngue (inglês, português), especialista, experiente no domínio dos esportes e nativo de um país de língua inglesa. Essa tradução *gold standard* foi utilizada como padrão de comparação de qualidade de TM, tendo passado por todo um processo de validação. Vejamos nas próximas seções a avaliação dos sistemas de TM através das métricas BLEU, TER e HTER.

6.2.1 Resultados de Avaliação BLEU dos Sistemas de TM S-Base, S-Pré e S-Pós

Os resultados de avaliação de TM seguindo a pontuação BLEU apresentados pelas traduções geradas pelos três sistemas de TM podem ser consultados na Tabela 5.

Tabela 5 – Avaliação de TM BLEU dos Sistemas S-Base, S-Pré e S-Pós

Sistema de TM	**S-Base**	**S-Pré**	**S-Pós**
BLEU	53.13	48.12	53.66

Compilado pelo autor (2020).

A partir dos dados exibidos na Tabela 5 acerca da avaliação de TM utilizando a métrica BLEU, percebemos que o sistema de TM estado da arte Google Tradutor apresentou uma pontuação de 53.13. O sistema de TM semanticamente enriquecido com injeção terminológica no pré-processamento gerou um score de 48.12. Por último, o segundo sistema de TM melhorado semanticamente, mas com injeção terminológica na etapa de pós-edição obteve uma pontuação BLEU de 53.66. Segundo os critérios de interpretação da BLEU expostos na Figura 48, quanto maiores são os *scores*, melhores são as traduções geradas. Assim, tanto o S-Base quanto o S-Pós geram traduções com alta qualidade, sendo consideradas adequadas e fluentes. Já o sistema de TM S-Pré tem suas traduções consideradas de alta qualidade. Na seção 6.2.4 discutiremos os *scores* gerados pela BLEU e uma interpretação comparada mais detalhada.

6.2.2 Resultados de Avaliação TER dos sistemas de TM S-Base, S-Pré e S-Pós

Os resultados de avaliação de TM com os *scores* da TER para as traduções geradas pelos três sistemas de TM estão expostos na Tabela 6.

Tabela 6 – Avaliação de TM TER dos Sistemas S-Base, S-Pré e S-Pós

Sistema de TM	**S-Base**	**S-Pré**	**S-Pós**
TER	36.23	42.63	36.47

Compilado pelo autor (2020).

A Tabela 6 exibe a avaliação de TM proposta seguindo a métrica TER. Essa métrica postula que quanto maior for o *score* de avaliação, maior será o esforço necessário por um humano para editar uma tradução avaliada para que ela se corresponda ao *gold standard*. Partindo disso, quanto menor for o *score* apresentado pela TER para as traduções avaliadas, melhores serão as traduções em comparação a outras. Nota-se novamente que o sistema de TM estado da arte Google Tradutor e o enriquecido semanticamente com injeção terminológica na pós-edição apresentam um melhor desempenho. O S-Base apresentou uma pontuação de 36.23 e o S-Pós gerou um score de 36.47. Ainda seguindo a avaliação pela TER, o sistema semanticamente melhorado com injeção terminológica no pré-processamento obteve uma pontuação de 42.63. Na seção 6.2.4 discutiremos os valores TER comparados entre os sistemas de TM analisados nesta tese.

6.2.3 Resultados de Avaliação HTER dos sistemas de TM S-Base, S-Pré e S-Pós

A pontuação HTER de avaliação de TM para as traduções geradas pelos três sistemas de TM encontra-se ilustrada na Tabela 7.

Tabela 7 – Avaliação de TM HTER dos Sistemas S-Base, S-Pré e S-Pós

Sistema de TM	**S-Base**	**S-Pré**	**S-Pós**
HTER	13.80	10.44	7.38

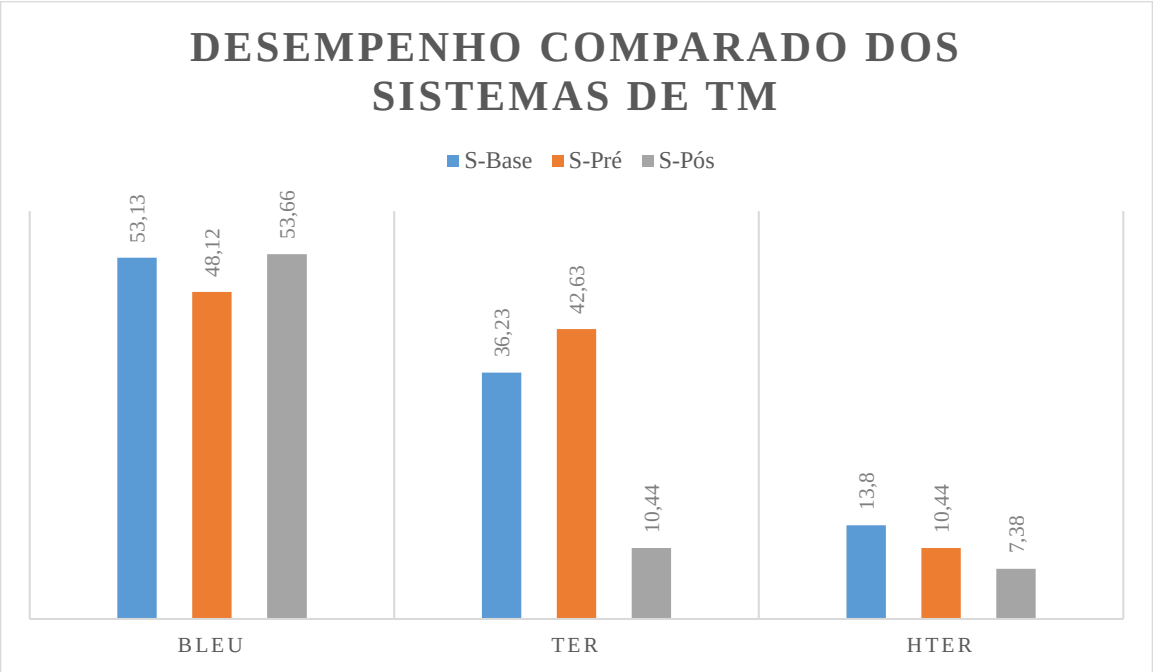
Compilado pelo autor (2020).

A partir da exposição da Tabela 7, constata-se os scores de avaliação HTER para os três sistemas aqui comparados, S-Base, S-Pré e S-Pós. O sistema de TM estado da arte Google Tradutor apresentou uma pontuação HTER de 13.80. O sistema de TM semanticamente enriquecido com injeção terminológica no pré-processamento gerou um *score* de 10.44. Por fim, o segundo sistema de TM melhorado semanticamente, mas com injeção terminológica na etapa de pós-edição obteve uma pontuação HTER de 7.38. Da mesma forma que a TER, quanto menor for o *score* da HTER gerado para uma dada tradução, melhor será sua avaliação em termos de qualidade e correspondência com o *gold standard*, dado o número menor de edições que foram realizadas. Na seção 6.2.4 discutiremos a pontuação HTER e uma interpretação comparada mais detalhada entre os três sistemas de TM aqui expostos.

6.2.4 Discussão dos Resultados

Ao longo deste trabalho foram desenvolvidos dois sistemas de TM semanticamente enriquecidos com frames e relações qualia. O primeiro possui a injeção terminológica realizada na etapa de pré-processamento e o segundo no estágio da pós-edição. Partindo-se disso, estabelece-se uma comparação, a partir de um mesmo corpus de sentenças de teste, entre as traduções geradas por ambos os sistemas de TM em contraste com traduções geradas por um sistema de TM estado da arte amplamente utilizado e reconhecido, o Google Tradutor. Na tentativa de avaliar os sistemas de TM, 3 métricas foram selecionadas para avaliação (BLEU, TER e HTER). Os dados gerados encontram-se esboçados no Gráfico 2 abaixo, sendo seguido por uma discussão comparativa para os três sistemas de TM (S-Base, S-Pré e S-Pós).

Gráfico 2 – Avaliação comparada dos Sistemas de TM pelas Métricas BLEU, TER e HTER



Elaborado pelo autor (2020).

A métrica BLEU avalia a correspondência de ocorrências de n-grams entre as traduções geradas e o *gold standard*, posições das palavras nas sentenças e não possui um tratamento específico voltado à semântica em contextos mais específicos.

Seguindo a avaliação da BLEU para os três sistemas, o S-Pré mostra-se como sendo um sistema de TM inferior aos outros dois (S-Base e S-Pré), provavelmente por apresentar alguns problemas nos aspectos sintáticos da tradução como trocas nas flexões de singular e plural, troca nos artigos definidos e indefinidos, concordância verbal, a colocação de preposições inadequadas para certos verbos, e ainda, um ou outro problema de tradução lexical como é o caso da sentença 43, em que "equipamento de segurança" foi traduzido como "*safety equipment*" (S-Base) e "*security guard apparatus*" (S-Pré). Entretanto, a métrica BLEU, por não apresentar um tratamento semântico adequado em seu processo de avaliação, desconsidera todas as traduções corretas dentro do domínio específico dos esportes geradas pelo sistema S-Pré, colocando-o como inferior. Todavia, mesmo apresentando certos problemas sintáticos, ela é semanticamente superior ao sistema S-Base, fato esse que será ilustrado nos resultados oferecidos pela métrica HTER. Para um detalhamento maior acerca do funcionamento do sistema S-Pré, consultar o estudo preliminar realizado e ilustrado no Apêndice A.

Ainda segundo a pontuação BLEU, o sistema S-Base é avaliado como sendo ligeiramente inferior ao S-Pós na geração de traduções. O primeiro aspecto que influencia no

alto score de avaliação da BLEU para ambos os sistemas se coloca no fato de os dois sistemas possuírem um tratamento adequado para os aspectos sintáticos. Entretanto, há problemas semânticos lexicais de ambiguidade, específicos no domínio dos Esportes em algumas sentenças (2, 5, 11, 12, 15, 17, 20, 29, 31, 32, 38, 39, 44, 47 e 49), de que o sistema S-Base não consegue dar conta, enquanto que o Sistema S-Pós resolve a maioria dos casos, apresentando problemas apenas na tradução de "zona de lançamento.n, lançamento.n, lance.n e abertura.n" do total de 15 sentenças que continham palavras polissêmicas dos esportes.

Com isso, podemos afirmar que a métrica BLEU não se mostra eficaz ao lidar com a semântica das línguas, trabalha com corpora de treinamento de domínio genérico, foca mais na posição e correspondência formal entre as traduções e o *gold standard*, desconsiderando aspectos semânticos como, por exemplo, a seleção adequada de equivalentes de tradução para palavras polissêmicas em domínios específicos.

A métrica de avaliação de tradução por máquina TER calcula o número mínimo de edições necessárias para que uma hipótese (tradução que está sendo avaliada) seja exatamente uma das referências (podendo haver um ou mais textos traduzidos de referência). A normalização da TER é feita através da divisão pelo número médio de palavras das referências (havendo apenas uma referência, o número de palavras daquela referência). A TER não envolve humanos no processo de edições e faz o cálculo tendo uma tradução analisada e uma tradução de referência, acabando por priorizar o aspecto formal na avaliação das traduções. Nessa métrica de avaliação de TM, a pontuação é caracterizada de forma inversa, sendo que, quanto maior for o score TER, pior é a tradução, visto que muito esforço seria necessário para transformar a tradução avaliada na tradução de referência.

Partindo desse fato, o sistema de TM S-Base e S-Pós aparecem novamente como sendo superiores na avaliação das traduções, apresentando como *scores* 36.23 e 36.47, respectivamente. Os problemas de sintaxe apontados anteriormente para o sistema S-Pré podem ser indícios para o fato dele ter sido avaliado como uma pontuação TER superior de 42.43, indicando que um maior esforço seria feito em suas traduções para equivalerem ao *gold standard*. Há uma outra versão da TER, TERp, que considera sinônimos e paráfrases no processo. Entretanto, a TER se mostra também como uma métrica que avalia mais os aspectos formais, não considerando a semântica presente no texto traduzido.

Por questões de um alto custo no processo de implementação e o pouco tempo disponível, optou-se pela não utilização da métrica F-SEM, que, embora considere a semântica e seja uma boa opção para avaliação de tradução em projetos futuros, torna-se impraticável por

questões de seleção de frames e domínio específico, e ainda requer uma natureza de corpora e estrutura de anotação que não cabe nesta tese.

Levando-se em conta todas as métricas propostas acima e utilizadas mais a título de comparação, optamos pela utilização da versão que envolve mais humanos no processo de avaliação, a HTER. Essa métrica pode apresentar ruído, no que diz respeito à subjetividade tradutória. Todavia, foram selecionados três editores humanos, havendo uma média das edições propostas por eles. O cálculo do número e dos tipos de edição também foram feitos de forma independente por duas pessoas, sendo verificados por uma terceira. A métrica de avaliação de tradução por máquina HTER se mostrou a mais adequada semanticamente por estarmos trabalhando justamente com a tradução no domínio específico dos esportes, fator que implica a polissemia das palavras e a necessidade de um refinamento semântico no aspecto avaliativo.

Nas avaliações de TM, o S-Base gerou um score de 13.80, ficando o S-Pré com uma pontuação de 10.44 e o S-Pós com 7.38. Utilizando a métrica HTER, notamos que há uma inversão entre os sistemas avaliados em termos de qualidade de tradução, ficando o S-Pós como o sistema com a avaliação mais alta, sendo seguido pelo S-Pré e, por último, o S-Base. A inversão entre S-Pré e S-Base em comparação com as outras métricas utilizadas se deve ao fato de os editores humanos envolvidos na HTER realizarem correções – geralmente morfossintáticas – mínimas nas traduções analisadas para que elas mantenham o significado proposto pelo *gold standard* e estejam gramaticalmente corretas e compreensíveis na língua. **Os problemas lexicais de tradução em domínio específico apontados pelo S-Base mostram que em termos semânticos e de contexto, os dois sistemas aqui desenvolvidos geram melhores equivalentes de tradução para domínio específico**. A diferença significativa entre os dados ilustrados pela TER e a HTER também se devem ao componente humano envolvido na HTER, uma vez que os editores analisam as traduções como sendo boas ou adequadas independentemente do fato de elas possuírem as mesmas escolhas tradutórias da tradução de referência. A partir dos dados de avaliação fornecidos pela HTER, concluímos que as traduções geradas para o domínio específico pelo S-Pré e S-Pós correlacionam-se melhor com a tradução humana de referência do que traduções realizadas por sistemas de RNN que não levam em conta etapas semânticas em seu funcionamento. **Por fim, o sistema de TM semanticamente enriquecido com frames e relações qualia e com injeção terminológica na pós-edição demonstrou-se com um desempenho 50% superior ao sistema que representa o estado da arte**.

Concluindo, ressalta-se o destaque apontado a ambos os sistemas TM aqui desenvolvidos (S-Pré e S-Pós) por conseguirem gerar sentenças traduzidas no domínio

específico dos esportes com equivalentes de tradução adequados ao contexto específico, detectando questões ambíguas, tendo potencial de desambiguá-las através de um sistema de desambiguação (DAISY) e indicando à necessidade da utilização de métricas de avaliação de TM que não apenas considerem os aspectos formais no processo, mas os aspectos semânticos e contextuais.

7. CONCLUSÕES

O objetivo central desta tese foi desenvolver um sistema de tradução por máquina semanticamente enriquecido com frames e estrutura qualia que fosse capaz de gerar equivalentes de tradução adequados para domínios específicos, aqui, os esportes.

Com esse propósito, realizamos os seguintes passos de pesquisa: executamos a modelagem de frames e ULs para o domínio específico em português, inglês e espanhol. Estabelecemos relações de equivalência de tradução entre esses termos com base na consulta em manuais do esporte nas três línguas e documentos oficiais de associações esportivas. Modelamos também a estrutura qualia (formal, agentivo, constitutivo e télico) entre as ULs de domínio específico na base de dados da FrameNet Brasil como relações qualia ternárias mediadas por frames, tanto para o português, como para o inglês. Buscamos adquirir o conhecimento acerca dos tipos de algoritmos de tradução por máquina, passando por toda a teoria de TM, com o objetivo de descobrir um sistema em que componentes semânticos pudessem ser inseridos como etapa do algoritmo de TM. Em um estudo preliminar, desenvolvemos um sistema de TM semanticamente enriquecido com frames e qualia e injeção terminológica no pré-processamento. Detectamos alguns problemas sintáticos nesse sistema que semanticamente gera equivalentes de tradução adequados para o domínio específico. Partimos então para o desenvolvimento de outro sistema, também melhorado semanticamente com frames e qualia, mas com injeção terminológica na pós-edição. O passo seguinte foi a constituição de uma tradução humana de referência para que métricas pudessem ser empregadas na avaliação da qualidade das traduções geradas por ambos os sistemas aqui criados e um sistema de TM estado da arte. Por fim, utilizamos as métricas BLEU, TER e HTER na avaliação dos sistemas e percebemos a importância, além de sistemas de TM semanticamente melhorados, da existência de métricas de avaliação de TM que considerem também o aspecto semântico como critério de avaliação.

No capítulo 2, trouxemos os princípios da tradução por máquina, traçando um histórico dos sistemas de TM já desenvolvidos, resgatando características fundamentais de cada tipo de sistema, além de conceitos importantes para a área tais como as redes neurais, o pré-processamento, a pós-edição, a injeção terminológica e as métricas de avaliação de TM. No capítulo 3, abordamos as teorias linguísticas que sustentam esta tese, sejam elas a semântica de frames e a teoria do léxico gerativo, bem como o importante conceito de entidade. No capítulo 4, apresentamos o aplicativo m.knob (Multilingual Knowledge base), sendo um guia de bolso e intérprete pessoal enriquecido semanticamente. Trouxemos também toda a modelagem feita

na estrutura da base de conhecimento através de frames específicos do domínio dos esportes, relações qualia e o uso de ontologias. No capítulo 5, demonstramos metodologicamente como o corpus de sentenças dos esportes foi compilado, como se deu sua tradução para a língua-alvo por um tradutor humano especialista nativo de um país de língua inglesa, como ocorreu o processo de validação desse *gold standard*, além do funcionamento das métricas de tradução BLEU, TER e HTER dentro desta tese. Por fim, no capítulo 6, trouxemos os resultados gerados pelas métricas BLEU, TER e HTER para os três sistemas de TM avaliados (S-Base, S-Pré e S-Pós), além de discutirmos os impactos de tais medidas na avaliação de TM. Os dados detalhados do funcionamento das métricas foram expostos nos Apêndices C-K.

Nossa hipótese foi a de que equivalentes funcionais de tradução adequados ao domínio específico dos esportes e ao contexto das sentenças são gerados a partir do enriquecimento semântico (frames e relações qualia) de sistemas de TM que utilizam injeção terminológica nas etapas de pré-processamento e pós-edição. Houve a submissão de um corpus de 50 sentenças-teste aos sistemas de TM aqui desenvolvidos. **Nossa hipótese foi verificada a partir da avaliação das traduções através de métricas de avaliação de tradução (BLEU, TER e HTER), com o nosso sistema de TM com injeção terminológica na pós-edição apresentando um desempenho 50% melhor do que o sistema estado da arte (NMT Google Tradutor – V2), segundo a avaliação pela HTER**.

Ademais, os objetivos específicos foram alcançados nesta tese incluindo a modelagem do domínio dos esportes através da criação de frames e qualia, o desenvolvimento e teste de dois sistemas de TM enriquecidos semanticamente, estabelecimento e validação de uma tradução humana de referência, e o emprego de métricas de TM (BLEU, TER e HTER) na avaliação dos sistemas aqui desenvolvidos em comparação com o sistema estado da arte Google Tradutor.

As principais contribuições e inovações desta tese são:

- A modelagem do domínio específico dos Esportes (em três idiomas: português, inglês e espanhol) em termos de frames, unidades lexicais, relações entre frames, EF-frame e qualia.
- Proposição de um fluxograma de modelagem de domínios específicos com os passos necessários para a criação de um domínio específico qualquer.
- A criação de dois sistemas de TM híbridos, enriquecidos semanticamente com frames e qualia, que geram equivalentes de tradução para domínio específico mais adequados, a partir da injeção terminológica em etapas de pré-processamento e pós-edição.

Como desdobramentos e trabalhos futuros desta pesquisa estão:

- A inserção do sistema de TM semanticamente enriquecido com injeção terminológica na pós-edição dentro da aplicação m.knob para uso e testes de desempenho.
- A modelagem de outros domínios para a replicação de testes de um sistema de TM que apresenta um bom desempenho para domínios específicos.
- A modelagem e replicação de testes dentro do domínio específico dos esportes para outras línguas como o espanhol, italiano, francês, alemão, por exemplo.
- A continuidade das pesquisas nas áreas da linguística e da computação que envolvam a utilização da semântica de frames, estrutura qualia, sistemas de TM e métricas de avaliação de TM.
- A escrita de um dicionário de esportes multilíngue baseado em frames e relações qualia.
- A inserção ou combinação de *Knowledge Graph Embeddings* com os sistemas de TM aqui desenvolvidos com o propósito de aperfeiçoar a TM e o reconhecimento de ENs.

Essa tese gerou duas tecnologias na área de NLP, sendo os dois sistemas de TM semanticamente melhorados, além do aplicativo m.knob como um produto tecnológico. Os dois sistemas – S-Pré e S-Pós – tiveram seus pedidos de patente depositados junto ao INPI, sob a titularidade da Universidade Federal de Juiz de Fora e da FAPEMIG, que financiou parte da pesquisa relativa ao m.knob, e com autoria da equipe que atuou no desenvolvimento dos sistemas, composta pelo então bolsista de Iniciação em Desenvolvimento Tecnológico e Inovação, Mateus Coutinho Marim, pelo Dr. Ely Edison da Silva Matos, técnico do Laboratório FrameNet Brasil, pelo autor desta tese e por seu orientador, Dr. Tiago Timponi Torrent.

A produção acadêmica advinda desta pesquisa inclui esta tese em si, além de participações em congressos nacionais e internacionais na divulgação da ciência desenvolvida pelo laboratório da FrameNet Brasil. Como publicações resultantes desta pesquisa, elencamos:

- Capítulos de livros:

1. PERON-CORRÊA, S. R.; DINIZ, A.; TAVARES, T.; TORRENT, T. T. Contributions of frame semantics and construction grammar in the processes of machine translation. In: CUADRA, P. V.; REY, A. C.; GONZÁLEZ, P. C. (Orgs.). Nuevas tendencias en

traducción: Fraseología, Interpretación, TAV y sus didácticas. Berlim: Peter Lang, 2018, v. 1, p. 531-552.

2. PERON-CORRÊA, S. R.; DINIZ, A.; LARA, M.; MATOS, E.; TORRENT, T. FrameNet-Based Automatic Suggestion of Translation Equivalents. In: SILVA, J.; RIBEIRO, R.; QUARESMA, P.; ADAMI, A.; BRANCO, A. (Orgs.). Lecture Notes in Computer Science. Berlim: Springer International Publishing, 2016, p. 347-352.

- Trabalhos completos publicados em anais de eventos:

1. COSTA, A. D.; CZULO, O.; TORRENT, T. T.; MATOS, E. E. S.; KAR, D. Designing a Frame-Semantic Machine Translation Evaluation Metric. In: 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019), 2019, Varna, Bulgária. 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019), 2019. p. 28-35.
2. COSTA, A. D.; GAMONAL, M. A.; PAIVA, V. M. R. L.; MARCAO, N. D.; PERON-CORREA, S. R.; ALMEIDA, V. G.; MATOS, E. E. S.; TORRENT, T. T. FrameNet-Based Modeling of the Domains of Tourism and Sports for the Development of a Personal Travel Assistant Application. In: International FrameNet Workshop 2018: Multilingual FrameNets and Constructicons., 2018, Miyazaki. Proceedings of the LREC 2018 Workshop International FrameNet Workshop 2018: Multilingual Framenets and Constructicons. Paris: European Language Resources Association, 2018. v. 1. p. 6-12.
3. COSTA, Alexandre Diniz da; TORRENT, T. T. A Modelagem Computacional do Domínio dos Esportes na FrameNet Brasil. In: Symposium in Information and Human Language Technology, 2017, Uberlândia, MG. STIL 2017 XI Brazilian Symposium in Information and Human Language Technology and Collocated Events, 2017.

- Resumos publicados em Anais de Eventos:

1. COSTA, A. D.; ALMEIDA, V. G.; TORRENT, T. T. Improving the granularity of framenet-based lexical semantics - ternary qualia roles for the Sports domain in FrameNet Brasil. In: 8th International Biennial Conference of the French Association for Cognitive Linguistics (AFLiCo 8): Language, Cognition and Creativity de 05 a 07 de junho de 2019, 2019, Mulhouse, França. AFLiCo 8 | Langage, Cognition et Créativité, 2019. p. 20.

2. COSTA, A. D.; PAIVA, V. M. R. L.; PERON-CORREA, S. R.; TORRENT, T. T. Frame Semantics in the Development of Multilingual Lexical Resources. In: XI Congresso Internacional da Associação Espanhola de Linguística Cognitiva (AELCO), 2018, Córdoba, Espanha. XI AELCO: Research in Metonymy, Metaphor, Constructions and Frames: Scope, Models and Methods., 2018. p. 163.

Para concluirmos, ressaltamos a importância desta pesquisa para a área de tradução por máquina e o desenvolvimento de aplicações linguístico-computacionais que implementam aspectos semânticos, em especial, a semântica de frames e as relações qualia. Reforçamos a importância das pesquisas desenvolvidas nas universidades e as parcerias estabelecidas com outros pesquisadores, nacionais e internacionais, com o propósito de incentivar a divulgação científica e o desenvolvimento de novas pesquisas que expandem as áreas supracitadas.

REFERÊNCIAS

ALLEN, J. Post-editing. In: SOMERS, H. (Ed.). **Computers and Translation:** A translator's guide. Amsterdam, Países Baixos: John Benjamins Publishing Co., 2003, v. 35, p. 297-317.

ALMEIDA, V. G. **Identificação Automática de Construções de Estrutura Argumental**: Um Experimento a partir da Modelagem Linguístico-computacional das Construções Transitiva Direta Ativa, Ergativa e de Argumento Cindido. 2016. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2016.

ARNALDO, Antunes. **As Coisas**. 3ª. ed. São Paulo: Iluminuras, 1996.

BAHDANAU, D.; CHO, K.; BENGIO, Y. 2015. Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations (ICLR), 2015, San Diego, Estados Unidos. **Anais** […]. San Diego, Estados Unidos: ICLR, 2015, p.1-15.

BANERJEE, P.; NASKAR, S. K.; ROTURIER, J.: WAY, A.; VAN GENABITH, J. Domain Adaptation in SMT of User-Generated Forum Content Guided by OOV Word Reduction: Normalization and/or Supplementary Data?. In: 6º Annual Conference of the European Association for Machine Translation (EAMT), 6., 2012, Trento, Itália. **Anais** […]. Trento, Itália: EAMT, 2012, p. 169-176.

BANERJEE, S.; LAVIE, A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In: ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, 2005, Ann Arbor, Estados Unidos. **Anais** […]. Ann Arbor, Estados Unidos: ACL, 2005, p. 65-72.

BELCAVELLO, F.; VIRIDIANO, M.; COSTA, A. D.; MATOS, E. E.; TORRENT, T. T. Frame-Based Annotation of Multimodal Corpora: Tracking (A)Synchronies in Meaning Construction. In: LREC International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet. 2020, Marselha, França. **Anais** […]. Paris, França: European Language Resources Association (ELRA), 2020, p. 23-30.

BENTIVOGLI, L.; BISAZZA, A.; CETTOLO, M.; FEDERICO, M. Neural versus Phrase-Based Machine Translation Quality: a Case Study. In: 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP). Austin, Estados Univos. **Anais** […]. Nova Iorque, Estados Unidos: ACL, 2016, p. 257-267.

BOUILLON, P.; GASPAR, L.; GERLACH, J.; PORRO, V.; ROTURIER, J. Pre-editing by Forum Users: a Case Study. In: Workshop (W2) on Controlled Natural Language Simplifying Language Use - 9th International Conference on Language Resources and Evaluation (LREC)., n. 9, 2014, Reykjavik, Islândia. **Anais** […]. Reykjavik, Islândia: 9th International Conference on Language Resources and Evaluation (LREC), 2014, p. 3-10.

BRIDLE, B.; GILBERT, R.; PARRISH, M.; EKE, K.; PHIPPS, T. (Ed.). **The Sports Book: The Games-the Rules-the Tactics-the Techniques**. Nova Iorque: Dorling Kindersley Limited, 2011.

CHANDIOUX, J. METEO, An Operational System for the Translation of Public Weather Forecasts. In: FBIS Seminar on Machine Translation, n. 2, microfiche 46, 1976, Rosslyn, Estados Unidos, **Anais** […]. Rosslyn: AJCL, 1976, p. 27-36.

CHATTERJEE, R.; NEGRI, M.; TURCHI, M.; FEDERICO, M.; SPECIA, L.; BLAIN, F. Guiding Neural Machine Translation Decoding with External Knowledge. In: Conference on Machine Translation (WMT), v. 1: Research Papers, 2017, Copenhagen, Dinamarca. **Anais** […]. Copenhagen, Dinamarca: Association for Computational Linguistics, 2017, p. 157-168.

CHO, K.; MERRIENBOER, B.; BAHDANAU, D.; BENGIO, Y. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In: Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, n. 8, 2014, Doha, Catar. **Anais** […]. Doha, Catar: Association for Computational Linguistics, 2014, p. 103-111.

CORBEIL, J. C.; ARCHAMBAULT, A. **The Visual Dictionary of Sports and Games**. Montreal, Canadá: QA International, 2009.

COSTA-JUSSÁ, M. R.; FONOLLOSA, J. A. R. Latest trends in hybrid machine translation and its applications. In: COSTA-JUSSÁ, M. R.; FONOLLOSA, J. A. R. (Ed.). **Hybrid Machine Translation: integration of linguistics and statistics**, [s.l.]: Elsevier Ltd., Computer Speech and Language, v. 32, 2015, p. 3-10.

CRESTANI, F. Application of Spreading Activation Techniques in Information Retrieval. **Artificial Intelligence Review**, Chom, Suíça, v. 11, n. 6, p. 453–482, 1997.

CZULO, O. Aspects of a primacy of frame model of translation. In: HANSEN-SCHIRRA, S.; CZULO, O.; HOFMANN, S. **Empirical Modelling of Translation and Interpreting**. Berlim, Alemanha: Language Science Press, 2017. p. 465-490.

CZULO, O.; TORRENT, T. T.; MATOS, E. E. S.; COSTA, A. D; KAR, D. Designing a Frame-Semantic Machine Translation Evaluation Metric. In: 2nd Workshop on Human-Informed Translation and Interpreting Technology – HiT-IT, n. 2, 2019, Varna, Bulgária. **Anais** […]. Shoumen, Bulgária: Incoma Ltd., 2019, p. 28-35.

DELLAPIETRA, S.; DELLAPIETRA, V. Candide: A Statistical Machine Translation System. In: **Human Language Technology**, 1994, Plainsboro, Estados Unidos. **Anais** […]. Plainsboro, Estados Unidos: ACL, 1994, p. 8-11, 1994.

DOUGAL. D. K. **Improving the Quality of Neural Machine Translation Using Terminology Injection**. 2018. Dissertação (Mestrado em Ciências) – Departamento de Ciência da Computação, Universidade Brigham Young, Provo, Estados Unidos, 2018.

FILLMORE, C. J. An Alternative to Checklist Theories of Meaning. In: The First Annual Meeting of the Berkeley Linguistics Society, n. 1, 1975, Berkeley, Estados Unidos. **Anais** […]. Berkeley, Estados Unidos: The Society, 1975, p. 123-131.

FILLMORE, C. J. The Case for Case Reopened. In: COLE, P.; SADOCK, J. (Eds.). **Syntax and Semantics 8**: Grammar Relations. Nova Iorque, Estados Unidos: Academic Press, 1977a, v. 8, p. 59-81.

FILLMORE, C. J. Frame semantics. In: SEOUL INTERNATIONAL CONFERENCE ON LINGUISTICS, 1981, Seoul, Coréia do Sul. **Linguistics in the Morning Calm […]**. Seoul, Coréia do Sul: Hanshin Publishing Co. 1982, Seoul, Coreia do Sul: 1982, p. 111-137.

FILLMORE, C. J. Frames and the semantics of understanding. In: **Quaderni di Semantica**. Bolonha, Itália: Societa Editrice il Mulino, 1985, v. 6, n. 2, p. 222-254.

FILLMORE, C. J. Border Conflicts: FrameNet Meets Construction Grammar. In: XIII EURALEX, 2008, Barcelona, Espanha. **Anais** […]. Barcelona, Espanha: Documenta Universitaria, 2008, p. 49-68.

FILLMORE, C. J.; BAKER, C. A Frames Approach to Semantic Analysis. In: HEINE, B; HEIKO, N. (Eds.). **The Oxford Handbook of Linguistic Analysis**. Nova Iorque, Estados Unidos: The Oxford University Press, 2010, p. 313-339.

FORTIN, J. (Ed.). **Enciclopedia Visual de los Deportes**. Badalona, Espanha: Editorial Paidotribo, 2008.

FOX, P. Using Heuristics. In: SAL. **Khan Academy**. 2019. Disponível em: https://www.khanacademy.org/computing/ap-computer-science-principles/algorithms-101/solving-hard-problems/a/using-heuristics Acesso em: 28 set. 2020.

FREITAS, A.; BARRETO, M. **Almanaque Olímpico**: Especial Jogos Rio 2016. 3ed. Rio de Janeiro, Brasil: Casa da Palavra/Grupo LeYa, 2016.

GALETZKA, M. **Intelligent Predictions:** an Empirical Study of the Cortical Learning Algorithm. 2014. Dissertação (Mestrado em Ciência da Computação) – Departamento de Ciência da Computação, Universidade de Ciências Aplicadas de Mannheim, Mannheim, 2014.

GAMONAL, M. A. **Copa 2014 FrameNet Brasil**: Diretrizes para a Constituição de um Dicionário Eletrônico Trilíngue a partir da Análise de Frames da Experiência Turística. 2013. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

GEERAERTS, D.; CUYCKENS, H. Introducing Cognitive Linguistics. In: GEERAERTS, D.; CUYCKENS, H. **The Oxford Handbook of Cognitive Linguistics**. Nova Iorque, Estados Unidos: Oxford University Press, 2007. p. 3-21.

GERMANN, U. Dynamic Phrase Tables for Machine Translation in an Interactive Post-editing Scenario. In: Workshop on Interactive and Adaptive Machine Translation, 2014, Vancouver, Canadá. **Anais** […]. Vancouver, Estados Unidos: Association for Machine Translation in the Americas (AMTA), 2014, p. 20-31.

GOLDBERG, Y. **Neural Network Methods for Natural Language Processing**. Synthesis Lectures on Human Language Technologies. San Rafael, Estados Unidos: Morgan & Claypool Publishers, 2017.

GOMES, A. N. M. **Tradução Automática e Linguagens Controladas**: Contributos para um Português Controlado. Dissertação (Mestrado em Tradução) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 2010.

GOMES, D. S. **Frames do Turismo Esportivo no Dicionário COPA 2014_FrameNet Brasil**. 2014. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2014.

GOOGLE. In: Google API. 2020. Disponível em: https://developers.google.com/places/android-api/start Acesso em: 01 jul. 2020.

GOOGLE. Produtos de IA e Machine Learning – Avaliar Modelos. Disponível em: https://cloud.google.com/translate/automl/docs/evaluate#bleu Acesso em: 07 nov. 2020.

GOOGLE TRADUTOR. In: Google Tradutor. 2020. Disponível em: https://translate.google.com.br/ Acesso em: 12 out. 2020.

GOUTTE, C.; CANCEDDA, N.; DYMETMAN, M.; FOSTER, G. **Learning Machine Translation**. Cambridge, Estados Unidos: MIT Press, 2009.

GRUBER, T. Ontology. In: LIU, L.; ÖZSU, M. T. (Eds.). **Ontology in Encyclopedia of Database Systems**. Nova Iorque, Estados Unidos: Springer-Verlag, 2009, p. 1963-1965.

GURNEY, K. **An Introduction to Neural Networks**. Londres, Inglaterra: UCL Press, 1997.

HAN, L. Machine Translation Evaluation Resources and Methods: A Survey. In: IPRC-2018 – Ireland Postgraduate Research Conference, 2018, Dublin, Irlanda. **Anais** […]. Dublin, Irlanda: DCU Postgraduate Society, 2018. Disponível em: https://arxiv.org/abs/1605.04515 Acesso em: 03 out. 2020.

HARVEY, M. Traduire l'intraduisible: Stratégies d'équivalence dans la traduction juridique. **ILCEA**, v. 3, p. 39-49, 2002. Disponível em: http://journals.openedition.org/ilcea/790 Acesso em: 15 dez. 2020.

HAYKIN, S. **Neural networks and learning machines**. 3. ed., Nova Jersey, Estados Unidos: Prentice Hall, 2008.

HELBIG, H. **Knowledge Representation and the Semantics of Natural Language**. Berlim, Alemanha: Springer, 2006.

HOBSON A. **The Oxford Dictionary of Difficult Words**. 1. ed., Nova Iorque, Estados Unidos: Oxford University Press, 2001.

HURTADO ALBIR, A. A Aquisição da Competência Tradutória: Aspectos Teóricos e Didáticos. In: PAGANO, A.; MAGALHÃES, C.; ALVES, F. (Org.). **Competência em Tradução**: cognição e discurso. Belo Horizonte: Editora da UFMG, 2005, p. 19-57.

HUTCHINS, W. J.; SOMERS, H. L. **An Introduction to Machine Translation**. San Diego, Estados Unidos: Academic Press, 1992.

JUNCZYS-DOWMUNT, M.; GRUNDKIEWICZ, R. Log-linear Combinations of Monolingual and Bilingual Neural Machine Translation Models for Automatic Post-Editing. In: First Conference on Machine Translation (WMT16), 2016, Berlim, Alemanha. **Anais** […]. Berlim, Alemanha: ACL, 2016, p. 751–758.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing**: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. 2. ed. Nova Jersey: Prentice Hall. 2008.

KAHNEMAN, D.; TVERSKY, A. Judgment under Uncertainty: Heuristics and Biases. **Science**, Washington DC, Estados Unidos, v. 185, n. 4157, p. 1124-1131, 1974.

KAMRAN, A. **Hybrid Machine Translation**. 2013. Tese (Doutorado em Linguística Formal e Aplicada) – Faculdade de Matemática e Física, Universidade Charles, Praga, 2013.

KARLBOM, H. **Hybrid Machine Translation**: Choosing the best translation with Support Vector Machines. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Departamento de Tecnologia da Informação, Universidade de Uppsala, Uppsala, 2016.

KILGARRIFF, A.; RYCHLÝ, P.; SMRŽ, P.; TUGWELL, D. The Sketch Engine. In: 11th EURALEX, n. 11, 2004, Lorient, França. **Anais** […]. Lorient, França: EURALEX, 2004, p. 105-115.

KLEIN, G.; KIM, Y.; DENG, Y.; NGUYEN, V.; SENELLART, J.; RUSH, A. M. OpenNMT: Neural Machine Translation Toolkit. In: The 13th Conference of the Association for Machine Translation in the Americas, n. 13, 2018, Boston, Estados Unidos. **Anais** […]. Boston, Estados Unidos: AMTA, 2018, p. 177-184.

KOEHN, P. **Statistical Machine Translation**. Nova Iorque, Estados Unidos: Cambridge University Press, 2010.

KOEHN, P. **Neural Machine Translation**. Berkeley, Estados Unidos: Universidade Johns Hopkins, 2017. Disponível em: https://arxiv.org/abs/1709.07809 Acesso em: 28 set. 2020.

KOEHN, P.; KNOWLES, R. Six challenges for neural machine translation. In: First Workshop on Neural Machine Translation, 2017, Vancouver, Canadá. **Anais** […]. Vancouver, Canadá: ACL, 2017, p. 28-39.

KOK, D.; BROUWER, H. **Natural Language Processing for the Working Programmer**. [*s. l.*]: Openlibra, 2010.

LIBERMAN, M. HTER. 2008. Disponível em: https://languagelog.ldc.upenn.edu/nll/?p=193 Acesso em: 13 out. 2020.

LINGUATEC. In: Personal Translator Demo. 2020. Disponível em: https://www.linguatec.de/personal-translator-demo/ Acesso em: 12 out. 2020.

LO, C. K.; WU, D. MEANT: An inexpensive, high-accuracy, semi-automatic metric for

evaluating translation utility via semantic frames. In: The 49th Annual Meeting of the Association for Computational Linguistics, n. 49, 2011, Portland, Estados Unidos. **Anais** […]. Portland, Estados Unidos: ACL, p. 220-229, 2011.

LO, C. K.; BELOUCIF, M.; SAERS, M.; WU, D. XMEANT: Better Semantic MT Evaluation without Reference Translations. In: The 52nd Annual Meeting of the Association for Computational Linguistics, n. 52, 2014, Baltimore, Estados Unidos. **Anais** […]. Baltimore, Estados Unidos: ACL, 2014, p. 765-771.

LO, C. K. MEANT 2.0: Accurate semantic MT evaluation for any output language. In: The Conference on Machine Translation (WMT), 2017, Copenhagen, Dinamarca. **Anais** […]. Copenhagen, Dinamarca: ACL, 2017, p. 589-597.

LUONG, M. T.; PHAM, H.; MANNING, C. D. Effective Approaches to Attention-based Neural Machine Translation. In The 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015, Lisboa, Portugal. **Anais** […]. Lisboa, Portugal: ACL, 2015, p. 1412-1421.

MATOS, E. E. S. **LUDI**: Um Framework para Desambiguação Lexical com Base no Enriquecimento da Semântica de Frames. 2014. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2014.

MAULDIN, M. L. Chatterbots, Tinymuds, and the Turing Test: entering the Loebner Prize Competition. In: The Twelfth National Conference on Artificial Intelligence, n.12, 1994, Seattle, Estados Unidos. **Anais** […]. Seattle, Estados Unidos: AAAI Press, v. 1, p. 16-21, 1994.

MINSKY, M. A Framework for Representing Knowledge. Massachusetts, Estados Unidos: Massachusetts Institute of Technology Cambridge, 1974.

MOLCHANOV, A. PROMT DeepHybrid system for WMT12 shared translation task. In: 7th Workshop on Statistical Machine Translation, n. 7, 2012, Montreal, Canadá. **Anais** […]. Montreal, Canadá: ACL, 2012, p. 345-348.

MOLCHANOV, A.; BYKOV, F. PROMT Translation Systems for WMT 2016 Translation Tasks In: The First Conference on Machine Translation, v. 2, 2016, Berlim, Alemanha. **Anais** […]. Berlim, Alemanha: ACL, 2016, p. 339-343.

MORAVCSIK, J. M. Aitia as generative factor in Aristotle's philosophy. **Dialogue: Canadian Philosophical Review**. Cambridge, Inglaterra, v. 14, n. 4, p. 622-638, 1975.

NAGAO, Makoto. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In: The International NATO Symposium on Artificial and Human Intelligence, 1984, Amsterdam, Países Baixos. **Anais** […]. Nova Iorque, Estados Unidos: Elsevier North-Holland, 1984, p. 173-180.

NAVIGLI, R.; PONZETTO, S. P. BabelNet : The Automatic Construction, Evaluation and Application of a Wide-coverage Multilingual Semantic Network. **Artificial Intelligence**, Amsterdam, Países Baixos, v. 193, 2012, p. 217-250.

PAIVA, V. M. R. L. **Recomendação Automática de Atrações Turísticas a partir da Análise Semântica de Comentários de Usuários de Plataformas Colaborativas**: Uma Aplicação da FrameNet Brasil. 2019. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2019.

PAPIENI, K.; ROUKOS, S.; WARD, T.; ZHU, W. J. BLEU: a Method for Automatic Evaluation of Machine Translation. In: The 40th Annual Meeting of the Association for Computational Linguistics (ACL), n. 40, 2002, Filadélfia, Estados Unidos. **Anais** […]. Filadélfia, Estados Unidos: ACL, 2002, p. 311-318.

PERON-CORREA, S. R. **A Semântica de Frames na Constituição de Dicionários Temáticos Multilíngues para Usuários Não-Especialistas**: Interface, Interação e Avaliação. 2019. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2009.

PERON-CORRÊA, S. R.; DINIZ, A.; LARA, M.; MATOS, E.; TORRENT, T. FrameNet-Based Automatic Suggestion of Translation Equivalents. In: SILVA, J.; RIBEIRO, R.; QUARESMA, P.; ADAMI, A.; BRANCO, A. (Orgs.). **Lecture Notes in Computer Science**. Berlim: Springer International Publishing, 2016, p. 347-352.

POIBEAU, T. **Machine Translation**. Cambridge, Estados Unidos: MIT Press, 2017.

POPOVIC, M. CHRF++: words helping character n-grams. In: Conference on Machine Translation (WMT), Volume 2: Shared Task Papers, 2017, Copenhagen, Dinamarca. **Anais** […]. Copenhagen, Dinamarca: ACL, 2017, p. 612–618.

PRADHAN, S.; WARD, W.; HACIOGLU, K.; MARTIN, J. H.; JURAFSKY, D. Shallow Semantic Parsing using Support Vector Machines. In: Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL), 2004, Boston, Estados Unidos. **Anais** […]. Boston, Estados Unidos: ACL, 2004, p. 233-240.

PUSTEJOVSKY, J. **The Generative Lexicon**. Cambridge, Estados Unidos: MIT Press, 1995.

PUSTEJOVSKY, J.; STUBBS, A. **Natural Language Annotation for Machine Learning**: A guide to corpus building for applications. Sebastopol, Rússia: O'Reilly, 2012.

PUSTEJOVSKY, J.; JEZEK, E. Integrating Generative Lexicon and Lexical Semantic Resources. In: 10th Language Resources and Evaluation Conference, n. 10, 2016, Portorož, Eslovênia. **Anais** […]. Portorož, Eslovênia: LREC, 2016, p. 3-139.

REFERENCE STANDARD. In: FARLEX PARTNER MEDICAL DICTIONARY. [Feasterville, Estados Unidos: Farlex Inc., 2012] Disponível em: https://medical-dictionary.thefreedictionary.com/Reference+Standard Acesso em: 10 out. 2020.

RODRIGUES, D.; NUNO, F.; SALERNO, S. O Livro dos Esportes: os Esportes, as Regras, as Táticas, as Técnicas. Rio de Janeiro, Brasil: Nova Fronteira, 2012.

ROOM, A. **Dictionary of Sports and Games Terminology**. Londres, Inglaterra: McFarland & Company, 2010.

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M.; JOHNSON, C.; SCHEFFCZYK. **FrameNet II**: Extended theory and practice. Disponível em: http://framenet.icsi.berkeley.edu/. Acesso em: 03 out. 2020.

SALVA, C. A. C. **El Libro de los 1001 Porqués de los Deportes**. Madri, Espanha: Editorial Visor, 2011.

SHIMANAKA, H.; KAJIWARA, T.; KOMACHI, M. Machine Translation Evaluation with BERT Regressor. 2019. arXiv preprint arXiv:1907.12679.

SIMARD, M.; UEFFING, N.; ISABELLE, P.; KUHN, R. Rule-based Translation With Statistical Phrase-based Post-editing. In: The Second Workshop on Statistical Machine Translation, 2007, Praga, República Tcheca. **Anais** […]. Praga, República Tcheca: ACL, 2007, p. 203-206.

SNOVER, M. **Translation Error Rate**. Disponível em: http://www.cs.umd.edu/~snover/tercom/ Acesso em: 07 nov. 2020.

SNOVER, M.; DORR, B.; SCHARTZ, R.; MICCIULLA, L.; MAKHOUL, J. A Study of Translation Edit Rate with Targeted Human Annotation. In: Association for Machine Translation in the Americas, 2006, Boston, Estados Unidos. **Anais** […]. Boston, Estados Unidos: AMTA, 2006, p. 223-231.

SOUZA, B. C. P. **Frames de Turismo como Negócio no Dicionário COPA 2014_FrameNet Brasil**. 2014. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2014.

SPARAVIGNA, A. C.; MARAZZATO, R. **Using Google Ngram Viewer for Scientific Referencing and History of Science**. 2015. Disponível em: https://arxiv.org/abs/1512.01364 Acesso em: 20 set. 2019.

SUTSKEVER, I, VINYALS, O. LE, Q. V. Sequence to sequence learning with neural networks. In: 27th International Conference on Neural Information Processing Systems, 2014, Bangkok, Tailândia. **Anais** […] Cambidge, Estados Unidos: MIT Press, 2014, p. 3104-3112.

SUTTON, C.; MCCALLUM, A. An Introduction to Conditional Random Fields. In: **Machine Learning**. Boston: NOW Foundations and Trends, 2012, v. 4. n. 4, p. 267-373.

TILDE. Interactive BLEU Score Evaluator. Disponível em: https://www.letsmt.eu/Bleu.aspx Acesso em: 07 nov. 2020.

TORRENT, T. T.; SALOMÃO, M. M. M.; MATOS, E. E.; GAMONAL, M. A.; GONCALVES, J.; SOUZA, B. C. P.; GOMES, D. S.; PERON, S. R. Multilingual lexicographic annotation for domain-specific electronic dictionaries: The Copa 2014 FrameNet Brasil Project. **Constructions and Frames**, Amsterdam, Países Baixos, v. 6, p. 73-91, 2014b.

THOUIN, B. The MÉTÉO System. **Practical Experience of Machine Translation**, Amsterdam, Países Baixos, p. 39-44, 1982.

TRUJILLO, A. **Translation Engines**: Techniques for Machine Translation. Londres, Inglaterra: Springer-Verlag, 1999.

VAN HEES, M.; KOZŁOWSKA, P.; TIAN, N. Web-based automatic translation: the Yandex. Translate API. 2015. Disponível em: https://mediatechnology.leiden.edu/openaccess/web-technology-reports Acesso em: 28 set. 2020.

WICENTOWSKI, R.; KELLY, M.; LEE, R. SWAT: Cross-Lingual Lexical Substitution using Context Matching and Machine Translation. In: 5TH INTERNATIONAL WORKSHOP ON SEMANTIC EVALUATION, 2010, Uppsala, Suécia. **Anais** […]. Uppsala, Suécia: ACL, 2010, p. 123-128.

WU, Y.; SCHUSTER, M.; CHEN, Z., LE, Q. V.; NOROUZI, M.; MACHEREY, W.; KRIKUN, M.; Y. CAO, Y.; GAO, Q.; MACHEREY, K. *et al*. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. In: **Semanticscholar**. 2016. Disponível em: https://arxiv.org/pdf/1609.08144.pdf Acesso em: 03 out. 2020.

APÊNDICE A – AVALIAÇÃO PRELIMINAR DO DESEMPENHO DO SISTEMA DE TRADUÇÃO COM INJEÇÃO TERMINOLÓGICA NO PRÉ-PROCESSAMENTO

Em um estudo preliminar, foram selecionadas de manuais esportivos e sites de notícias 50 sentenças em língua portuguesa pertencentes ao domínio específico dos esportes que apresentavam palavras polissêmicas e continham relações qualia modeladas entre pelo menos duas ULs extraídas a partir de palavras das sentenças. Essas 50 sentenças na língua-fonte (português) foram submetidas a um sistema de TM comumente utilizado que aplica Redes Neurais Recorrentes (RNNs). Fundamentando-se na análise dessas 50 sentenças traduzidas para a língua-alvo (no caso, o inglês), 15 sentenças que não apresentavam os equivalentes de tradução mais adequados dentro do domínio específico dos esportes foram separadas. As mesmas 50 sentenças na língua-fonte (português) foram aplicadas no sistema de TM aqui desenvolvido que realiza a injeção terminológica na etapa de pré-processamento em sentenças intermediárias híbridas (português-inglês).

Posteriormente a isso, os resultados de tradução obtidos tanto pelo sistema de TM (por RNNs) comumente utilizado, quanto aqueles do sistema de TM semanticamente melhorado com injeção terminológica no pré-processamento passaram por um processo de avaliação das traduções. Foi solicitado a um grupo de 10 tradutores (alunos ou graduados do Bacharelado em Tradução em Letras: Inglês-Português) que avaliassem a adequação de ambas as traduções em relação ao domínio específico dos esportes em um formulário. Da composição desse grupo de 10 tradutores, 9 já eram graduados (90%) e 1 (10%) estava no sétimo período do curso. No que concerne ao nível de inglês dos avaliadores, 8 avaliadores (80%) julgaram possuir um nível de inglês excelente, e 2 tradutores (20%) se posicionaram como possuindo um nível de inglês bom.

Passando ao processo de avaliação em si, foi disponibilizado ao tradutor a sentença original em português. Logo após, foram mostradas as duas traduções daquela sentença em inglês, sendo uma gerada pelo sistema de TM por RNNs simples e a outra pelo sistema de TM com injeção terminológica no pré-processamento. A ordem das sentenças traduzidas foi aleatória ao longo da avaliação. A partir daí, pediu-se que os tradutores avaliassem as duas traduções oferecidas em inglês para cada sentença do português em relação a sua adequação ao domínio específico dos Esportes. Essa avaliação deveria ser feita com a utilização de notas 1 a 5, sendo 1 para uma tradução totalmente inadequada e 5 para uma tradução totalmente adequada. Foi também informado ao avaliador que a avaliação de ambas as traduções para cada sentença era independente, podendo ou não receberem a mesma avaliação por se tratar de traduções diferentes. Considerando todas as avaliações de cada tradução, foi computado um

score médio de avaliação. A Tabela 8 apresenta a compilação realizada e ilustra em suas colunas: a sentença na língua-fonte (português), a sentença traduzida na língua-alvo (inglês) por um sistema de RNNs, o score médio de avaliações dessa tradução, a sentença traduzida na língua-alvo (inglês) pelo sistema de TM proposto no âmbito desta tese, enriquecido semanticamente com frames e relações qualia, sendo a injeção terminológica realizada na etapa de pré-processamento, e por último, o score médio de avaliação dessas traduções.

Tabela 8 – Resultados do experimento de avaliação de TM por tradutores

	Sentença (PT)	**Sentença traduzida (EN) por um sistema de TM por RNN (S-Base)**	**Score médio de avaliação da TM (RNN)**	**Sentença traduzida (EN) por um sistema de TM por RNN semanticamente enriquecido com frames e qualia, além da injeção terminológica no pré-processamento (S-Pré)**	**Score médio de avaliação da TM (RNN enriquecido e com injeção terminológica no pré-processamento)**
1	Um corredor não tenta correr uma maratona nos primeiros dias de treinos.	A runner does not try to run a marathon in the first few days of training.	4,7	A racer does not try to run a marathon during the first days of training.	3,1
2	O lançador é desclassificado se sair da zona de lançamento antes, durante ou depois do lançamento.	The pitcher is disqualified if he leaves the launch zone before, during or after the launch.	3,8	The thrower is disqualified if he leaves the throwing zone before, during or after the throwing.	2,5
3	O árbitro Mário Yamasaki decidiu interromper a luta achando que o lutador havia desmaiado.	Referee Mario Yamasaki decided to stop the fight thinking that the fighter had passed out.	5,0	Judge Mário Yamasaki decided to stop the combat thinking the fighter had passed out.	1,6
4	Coloco o ponto em que o levantador executa o levantamento.	I place the point at which the lifter performs the lift.	4,6	I place the point at which the lifter performs the facelift.	1,5
5	O ponta é o jogador que menos tempo tem para pensar na armação de uma jogada.	The forward is the player who has less time to think about setting up a move.	3,2	The wing is the player that has less time to think in the setup of a play.	3,3
6	O ginásio possui uma quadra que pode receber jogos de futsal e handebol.	The gym has a court that can host futsal and handball games.	4,9	The gym has a court that can host a game of futsal and handball.	2,1
7	O OG Kyle Long sofreu uma lesão na mão durante o primeiro quarto do jogo contra os Saints.	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	4,5	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	4,5
8	Além disso, a rede possui 1,55 cm sendo mais alta que a utilizada no tênis e menor que a rede da quadra de vôlei.	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net on the volleyball court.	3,3	In addition, the net has 1.55 cm being higher than that used in tennis and smaller than the net of a volleyball court.	2,8
9	A competição de saltos em equipe envolve um saltador e uma saltadora.	The team jumping competition involves a jumper and a jumper.	1,8	The team jump dispute involves a vaulter and a vaulter.	2,8

10	O tênis, um dos esportes mais tradicionais e praticados no mundo.	Tennis, one of the most traditional and practiced sports in the world.	3,8	Tennis, one of the most traditional and practiced sports in the world.	3,8
11	A bandeja é quando o jogador faz a cesta bem próxima do aro.	The tray is when the player makes the basket very close to the ring.	1,2	The layup is when the player makes the basket very close to the hoop.	4,3
12	Durante uma jogada, um jogador pode dar até dois toques não consecutivos, de modo que a equipe só pode dar no total três toques na bola.	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	4,0	During a play, a player can give up to two non-consecutive touches, so the team can only give a total of three touches on a ball.	2,3
13	Mario Suárez chuta rente à trave do gol de Schwarzer.	Mario Suárez kicks close to the goal post of Schwarzer.	3,6	Mario Suárez kicks close to the post of Schwarzer's goal.	3,7
14	A vela é um esporte olímpico desde 1900.	Sailing has been an Olympic sport since 1900.	4,9	Sailing has been an olympic sport since 1900.	4,0
15	No último lance do jogo, o zagueiro Gustavo Gómez deu um carrinho no atacante corinthiano Jô e fez o pênalti.	In the last game of the game, defender Gustavo Gómez gave the striker corinthiano Jô a trolley and made the penalty.	1,0	In the last move of the game, center back Gustavo Gómez tackled the forward corinthian Jô and made the penalty.	3,9
16	O artilheiro é o jogador Evandro, que marcou sete gols.	The top scorer is the player Evandro, who scored seven goals.	3,7	The leading scorer is player Evandro, who scored seven goals.	3,5
17	Fuga é uma técnica utilizada no ciclismo de estrada.	Escape is a technique used in road cycling.	1,7	Breakaway is a technique used in road cycling.	4,7
18	A partida se inicia sempre com um saque, jogada que, obrigatoriamente, alterna-se entre os participantes a cada game.	The game always starts with a serve, a move that must alternate between the participants in each game.	4,2	The game always starts with a service, play that, obligatorily, alternates between the participants in each game.	2,4
19	Jogando na posição 3, um ala é o jogador que mais se aproxima dos dois extremos das posições do basquete.	Playing in position 3, a winger is the player who comes closest to the two ends of the basketball positions.	3,6	Playing in position 3, a winger is the player that most closely matches the two extremes of the position of the basketball.	2,4
20	A equipe jogava bem e Juciely, com uma china, jogada muito utilizada pela atleta, fez 09-06.	The team played well and Juciely, with a china, played very often by the athlete, made 09-06.	2,6	The team played well and Juciely, with a slide, play widely used by the sportsman, made 09-06.	2,9
21	No jogo de volta da decisão, torcedores do River apedrejaram o ônibus dos jogadores do Boca Juniors, que não tiveram condições de entrar no campo do Estádio Monumental de Nuñez, em Buenos Aires.	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	4,3	In the return game of the decision, supporter of the River stoned the bus of the players of the Boca Juniors, who had no condition to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	2,3
22	Com 18 anos, o capitão do time foi o jogador mais jovem da história do clube a marcar em um Atletiba.	At 18, the team captain was the youngest player in the club's history to score in an Atletiba.	4,8	At 18, the captain of the club was the youngest player in the history of the club to score in an Atletiba.	2,6

23	Então a piscina tem de ser dividida em dez raias para que só as oito internas, menos turbulentas, sejam usadas nas provas.	Then the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	1,9	Then the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the event.	4,1
24	O teoricamente dono da posição está novamente sem atuar, desta vez por lesão, enquanto o reserva é o jogador que mais vezes atuou nesta temporada.	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	4,2	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	4,3
25	O cruzamento é uma jogada forte nossa, como é de todas as equipes.	The cross is a strong move for us, as it is for all teams.	4,5	The cross is a strong play for us, as it is for all teams.	2,9
26	No entanto, a mídia esportiva costuma referir à jogada apenas como bicicleta, pouco empregando o prefixo chute ou pontapé.	However, the sports media usually refer to the play only as a bicycle, with little use of the prefix kick or kick.	2,4	However, the sports media usually refer to play only as a scissor kick, with little use of the prefix kick or kicking.	3,0
27	O estádio possui uma pista de atletismo de nove raias, dois telões gigantes e uma rede wi-fi de última geração.	The stadium has a nine lane athletics track, two giant screens and a state-of-the-art wi-fi network.	4,2	The stadium has a nine lane athletics track, two giant screens and a state-of-the-art wi-fi net.	3,5
28	Quatro países irão disputar o desafio dos quatro estilos da natação, onde borboleta é o nado mais complexo de aprender.	Four countries will contest the challenge of the four styles of swimming, where butterfly is the most complex swim to learn.	3,8	Four countries will compete for the challenge of the four stroke of a swimming, where fly is the most complex stroke to learn.	2,6
29	Segundo dados do Footstats, o ponta foi o jogador com mais cruzamentos certos e mais passes para gol na equipe.	According to data from the Footstats, the forward was the player with the most correct crosses and the most passes for goals in the team.	3,7	According to data from the footstats, the wing was the player with more certain crosses and more pass for goal en a team.	2,6
30	Totalmente diferente dos outros estilos, o nado peito exige muita coordenação e técnica do praticante.	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	4,5	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	4,5
31	Pedrinho dá um chapéu, mas não dá sequência na jogada, aos doze minutos do primeiro tempo.	Pedrinho gives a hat, but does not give sequence in the play, to the twelve minutes of the first half.	2,1	Pedrinho gives a lob, but does not give sequence in a play, to the twelve minutes of the first time.	3,5
32	A tradicional comemoração com um peixinho na quadra está viva na memória.	The traditional celebration with a goldfish on the court is alive in memory.	1,8	The traditional celebration with a dive en a court is alive in memory.	3,3
33	A posição e o alinhamento da gaiola no campo de competição é, portanto, crítico para o seu uso seguro.	The position and alignment of the cage in the competition field is therefore critical to its safe use.	4,5	The position and alignment of a cage in the dispute field is therefore critical to its safe use.	3,5
34	O nado de costas causa uma boa sensação após a execução de séries intensas de crawl, ou livre, e borboleta.	The backstroke causes a good sensation after the execution of intense series of crawl, or free, and butterfly.	3,8	The back stroke causes a good sensation after performing intense crawl, or free, and butterfly routines.	3,5

35	Muitos diziam que este era um salto criado pela escola queniana de atletismo, mas me parece que é uma variante do salto tesoura.	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	4,0	Many said that this was a jump created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissor jump.	4,3
36	Muitos acham que o gancho é o golpe onde o lutador lança sua mão de baixo para cima.	Many think that the hook is the stroke where the fighter throws his hand from the bottom up.	3,4	Many think that the hook is the blow where the fighter throws his hand from the bottom up.	4,2
37	Companheiro de Messi no Barcelona, o meia é o jogador brasileiro com o maior valor a atuar na Copa América.	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the Copa America.	4,4	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the America Cup.	4,1
38	Cinco séries de golpes combinados de suple, bombeiro e estabilização no solo com ênfase na precisão de movimento.	Five series of combined strokes of suple, fireman and ground stabilization with an emphasis on precision of movement.	3,6	Five combined stroke routines of suplex, fireman's carry and hold on the floor with an emphasis on movement accuracy.	4,2
39	No rugby, o abertura é o jogador mais habilidoso do time.	In rugby, the aperture is the most skilled player on the team.	2,0	In rugby, the fly-half is the most skilled player in the club.	4,3
40	O servidor é o jogador que coloca a bola em jogo para o primeiro ponto.	The server is the player who puts the ball in play for the first point.	4,2	The server is the player that puts the ball in game for the first point.	3,8
41	O sonho de Tristan Garcia, de 14 anos e fã de basquetebol era arremessar uma bola na cesta da quadra da escola.	The dream of Tristan Garcia, age 14 and a basketball fan, was to throw a ball into the basket on the school court.	4,6	The dream of Tristan Garcia, aged 14 and a basketball fan, was to throw a ball into a basket from a school court.	2,9
42	Um arco é um equipamento individual e pessoal.	A bow is individual and personal equipment.	4,7	A hoop is an individual and personal apparatus.	2,3
43	Invista em uma boa luva, a luva é o equipamento de segurança maior do boxe, ela pode evitar que você se machuque gravemente.	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	4,7	Invest in a good glove, the glove is the biggest security guard apparatus of boxing, it can prevent you from getting seriously hurt.	2,5
44	Todos os ginastas que disputam a prova saltam sobre um aparelho ligeiramente inclinado chamado mesa.	All gymnasts competing in the competition jump on a slightly inclined device called a table.	3,0	All gymnast competing in the event jump on a slightly inclined apparatus called a vault table.	4,3
45	Se a ginástica rítmica é a ovelha negra da família ginástica, em seguida, a corda é a ovelha negra dos aparelhos.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	4,3	If the rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	3,5
46	A fita é considerada o aparelho mais plástico e característico da ginástica.	The ribbon is considered the most plastic and characteristic apparatus of gymnastics.	4,7	The ribbon is considered the most plastic and characteristic apparatus of a gymnastics.	2,6
47	Ela venceu as disputas nos aparelhos: fita e maças.	She won the disputes on the devices: ribbon and apples.	1,5	She won the contest on the apparatus: ribbon and club.	4,4

48	A vara para salto é um equipamento muito avançado.	The jumping pole is very advanced equipment.	3,6	The pole for jump is a very advanced apparatus.	3,4
49	O atacante ainda protagonizou um lance de brilho ao aplicar um lençol em um adversário, jogada que chamou atenção de outros boleiros na web.	The striker also made a brilliant move by applying a sheet to an opponent, a move that drew the attention of other players on the web.	3,2	The forward also featured a brilliant move to apply a lob to an adversary, a play that drew the attention of other players on the web.	3,5
50	A prancha é o equipamento mais importante para pegar ondas grandes.	The board is the most important equipment for catching big waves.	4,2	The board is the most important apparatus for catching large waves.	3,0

Fonte: Elaborado pelo autor (2020).

A análise dos dados apresentados na Tabela 8 indica que, calculando-se um score médio das avaliações para as 50 traduções oferecidas pelo sistema convencional de TM por RNNs (S-Base), obtivemos um score médio de 3,61 em 5. Já para as traduções oferecidas pelo Sistema 2, o score médio para as 50 avaliações foi de 3,31. Ao observarmos a diferença de avaliação entre as traduções de ambos os sistemas, elencamos alguns problemas e situações que podem ter influenciado os tradutores a atribuírem uma classificação menor para certas traduções.

O primeiro ponto diz respeito aos aspectos lexicais específicos do domínio dos esportes que os avaliadores podem não ter tido tanta familiaridade. Por exemplo, a sentença 3 contém a palavra **árbitro.n** que pode ser traduzida de diversas formas dentro dos esportes tais como ***ref*.n**, ***official*.n**, ***referee*.n**, ***judge*.n**, entre outros. Por ***referee*.n** (S-Base) ser mais comum e frequente, dado o fato de essa ser a equivalência oficial usada no futebol, esporte mundialmente famoso, isso pode ter influenciado os avaliadores a atribuírem um score relativamente menor para a tradução contendo o equivalente ***judge*.n** (S-Pré). Outros exemplos análogos em que pode ter acontecido esse mesmo processo avaliativo ocorrem nas sentenças 42 e 12. Nas traduções da sentença 12, ***rally*.n** (S-Base) e ***hit*.n** (S-Base) são avaliadas com um melhor score, ficando ***play*.n** (S-Pré) e ***touch*.n** (S-Pré) com uma avaliação menor. Embora essas palavras tenham tido um score de avaliação inferior, ***play*.n** e ***touch*.n** são equivalentes de tradução possíveis nos esportes para **jogada.n** e **toque.n**, respectivamente. Outro caso com a mesma situação ocorre na sentença 42 em que a tradução de **arco.n** sendo ***bow*.n** (S-Base) possui uma avaliação substancialmente maior do que a que contém ***hoop*.n** (S-Pré). Ao analisarmos especificamente as traduções, constatamos que ambas são traduções possíveis para **arco.n**, sendo ***bow*.n** no tiro com arco, e ***hoop*.n** na ginástica rítmica, fato este que pode ter passado despercebido pelos avaliadores.

O segundo ponto refere-se a duas situações específicas de geração de equivalentes de tradução inadequados. O primeiro exemplo encontra-se na sentença 4 com a tradução de

levantamento.n como ***lift*.n** (S-Base) e ***facelift*.n** (S-Pré). Esse problema pode ter ocorrido pela injeção de ***lift*** na sentença intermediária híbrida, o que fez o sistema reconhecer essa palavra dentro da língua portuguesa como um procedimento estético e traduzi-lo para o inglês. O mesmo acontece na sentença 43 em que "equipamento de segurança" foi traduzido como ***safety equipment*** (S-Base) e ***security guard apparatus***" (S-Pré). Nota-se o reconhecimento da polissemia da palavra "**segurança**", tendo sido traduzida corretamente pelo S-Base e como o profissional da área de segurança pelo S-Pré.

O terceiro ponto foi a flexão de plural ausente em algumas traduções do S-Pré. Essa situação pode ser verificada nas traduções das sentenças 6 (**jogos**, S-Base: ***games***, S-Pré: ***game***), 19 (posições, S-Base: ***positions***, S-Pré: ***position***), 21 (torcedores, S-Base: ***fans***, S-Pré: ***supporter***), 28 (estilos, S-Base: ***styles***, S-Pré: ***stroke***), 29 (passes, S-Base: ***passes***, S-Pré: ***pass***) e 44 (ginastas, S-Base: ***gymnasts***, S-Pré: ***gymnast***).

O quarto ponto corresponde à tradução de preposições inadequadas em um determinado contexto. Na sentença 5, temos o verbo pensar em (S-Base: ***think about***, S-Pré: ***think in***). O verbo ***think***, em inglês exige as preposições ***about*** ou ***of***, podendo existir a preposição ***in*** apenas em um contexto muito específico de "se pensar em uma língua". Por exemplo, "pensar em português (***think in Portuguese***). Portanto, essa inadequação preposicional é um fato que influencia na avaliação da tradução.

O quinto ponto relaciona-se à tradução de artigos definidos e indefinidos. Há três tipos de situação envolvendo os artigos. O apagamento do artigo é observado nas sentenças 16 (é o jogador, S-Base: ***is the player***, S-Pré: ***is player***), 18 (jogada que, S-Base: ***a move that***, S-Pré: ***play that***) e 20 (jogada muito utilizada pela, S-Base: ***played very often by the athlete***, S-Pré: ***play widely used by***). O artigo definido em português passando a indefinido em inglês devido ao artigo definido feminino singular em português "A" ser mantido na sentença intermediária injetada, o sistema passa a reconhecer como o artigo indefinido do inglês "A/AN". Consequentemente, a tradução gerada é de uma expressão indefinida em inglês, o que em português era uma expressão definida. Essa situação ocorre nas sentenças 29 (na equipe, S-Base: ***in the team***, S-Pré: ***en a team***), 31(na jogada, S-Base: ***in the play***, S-Pré: ***in a play***), 33 (da gaiola, S-Base: ***of the cage***, S-Pré: ***of a cage***), 41 (arremessar uma bola na cesta, S-Base: ***throw a ball into the basket***, S-Pré: ***throw a ball into a basket*** / da escola, S-Base:***on the school court***, S-Pré: ***from a school court***) e 46 (da ginástica, S-Base: ***of gymnastics***, S-Pré: ***of a gymnastics***). O último relacinado aos artigos é algo mais interno do S-Pré em que a expressão intermediária ***en*** foi incluída na tradução oficial gerada. Isso ocorreu na sentença 29 (na equipe,

S-Base: ***in the team***, S-Pré: ***en a team***) e 32 (na quadra, S-Base: ***on the court***, S-Pré: ***en a court***).

O sexto e último problema diz respeito à inadequação da estrutura sintática em dois momentos. Na sentença 29, há o uso de construção superlativa pelo S-Base e de construção comparativa pelo S-Pré. Entretanto, o mais adequado nesse contexto deveria ser o superlativo. O fragmento "o jogador com mais cruzamentos certos e mais passes para gol na equipe" foi traduzido como "***the player with the most correct crosses and the most passes for goals in the team***", pelo sistema 1, e como "***the player with more certain crosses and more pass for goal em a team"***. O outro problema sintático refere-se ao uso do verbo ***have*** (ter), em inglês, para se referir à altura, sendo que o correto seria utilizar o verbo ***be*** (ser/estar) nesse contexto. Essa situação ocorre na sentença 8 (a rede possui 1,55cm, S-Base: ***the net is 1.55cm***, S-Pré: ***the net has 1.55cm***).

Mesmo o S-Pré tendo apresentado os problemas descritos acima, há melhorias oferecidas no escopo de tradução lexical do domínio específico para os nomes de entidade nos Esportes. Inicialmente, ao analisarmos as 50 traduções em inglês fornecidas pelo Google tradutor (S-Base), separamos 15 sentenças (2, 5, 11, 12, 15, 17, 20, 29, 31, 32, 38, 39, 44, 47 e 49) em que as traduções de determinados nomes de entidade específicos dos esportes não foram adequadamente geradas pelo Sistema 1. Ao compararmos o score médio de avaliação dos tradutores para esse grupo de 15 sentenças, obtivemos um score médio de avaliação de 2,56 para as traduções do S-Base (Google Tradutor) e de 3,6 para as traduções do S-Pré (enriquecido semanticamente com frames e relações qualia, com injeção terminológica no pré-processamento). Observemos as equivalências de tradução para os nomes de entidade propostos por ambos os sistemas (S-Base e S-Pré) na Tabela 9.

Tabela 9 – Equivalentes de Tradução propostos (S-Base e S-Pré) para as 15 sentenças específicas

Número	**Termo Específico dos Esportes**	**Equivalente de Tradução (Sistema 1)**	**Equivalente de Tradução (Sistema 2)**	**Esporte**
2	lançador.n	pitcher.n	thrower.n	Atletismo
2	zona de lançamento.n	launch zone.n	throwing zone.n	Atletismo
2	lançamento.n	launch.n	throwing.n	Atletismo
5	ponta.n	forward.n	wing.n	Futebol
11	bandeja.n	tray.n	layup.n	Basquete
12	jogada.n	rally.n	play.n	Voleibol, Tênis, Badminton
15	deu um carrinho.v	gave a trolley.v	tackled.v	Futebol
17	fuga.n	escape.n	breakaway.n	Ciclismo
20	china.n	china.n	slide.n	Voleibol
29	ponta.n	forward.n	wing.n	Futebol

31	chapéu.n	hat.n	lob.n	Futebol
32	peixinho.n	goldfish.n.	dive.n	Voleibol
38	suple.n	suple.n	suplex.n	Luta Olímpica
38	bombeiro.n	fireman.n	fireman's carry.n	Luta Olímpica
38	estabilização.n	stabilization.n	hold.n	Luta Olímpica
39	abertura.n	aperture.n	fly-half.n	Rúgbi
44	aparelho	device.n	apparatus.n	Ginástica
47	maça.n	apple.n	club.n	Ginástica
49	lençol.n	sheet.n	lob.n	Futebol

Fonte: Elaborado pelo autor (2020).

Ao estabelecermos uma comparação entre os dois sistemas (S-Base e S-Pré) a partir da proposição de equivalentes de tradução para os nomes de entidade do domínio específico dos Esportes, nota-se que todos os equivalentes ilustrados na Tabela 9 foram gerados corretamente. Já os equivalentes propostos pelo sistema 1 até são possíveis para as palavras, mas em contextos diferentes dos esportes. Apenas a tradução de ***pitcher.n*** para **lançador.n** que também seria viável dentro do beisebol como uma das posições existentes. A modelagem de todos os termos específicos dos esportes e seus equivalentes foram modelados na base de dados da FN-Br a partir de manuais específicos dos esportes em três línguas (português, inglês e espanhol). A partir da diferença relativa no score médio de avaliação dos tradutores comparando essas 15 sentenças traduzidas por S-Base e S-Pré, temos uma corroboração de que o S-Pré justificar a oferece de fato uma melhoria semântica das traduções de domínio específico. Entretanto, na tentativa de elaborar um sistema de TM semanticamente enriquecido (com frames, qualia e um sistema de desambiguação de frames) que apresente menos problemas e, ainda sim, ofereça traduções de domínio específico adequadamente, justifica-se o desenvolvimento de um outro sistema com a injeção terminológica na etapa de pós-edição e a utilização de dicionários bilíngues na detecção de termos e equivalências. Esse outro sistema foi apresentado em 6.1.3, seus resultados ilustrados e discutidos em 6.2.

APÊNDICE B – VERIFICAÇÃO SEMÂNTICA DO GOLD STANDARD

Tabela 10 – Verificação de correspondência semântica de frames do domínio dos esportes evocados na base de dados pelas sentenças-fonte em português e as sentenças-alvo traduzidas para o inglês (*gold standard*)

Nº	Sentença-fonte (PT)	Frames dos Esportes Evocados	Sentença-alvo Gold Standard (EN)	Frames dos Esportes Evocados	Porcentagem de Correspondência
1	Um **corredor** não tenta **correr** uma **maratona** nos primeiros dias de treinos.	corredor.n: [frm_athletes_by_sport] correr.v: [frm_individual_moves] maratona.n: [frm_sports_disciplines]	A **runner** does not try to **run** a **marathon** in the first days of training.	runner.n: [frm_athletes_by_sport] run.v: [frm_individual_moves] marathon.n: [frm_sports_disciplines]	$\frac{3}{3} = 1$
2	O **lançador** é **desclassificado** se sair da **zona de lançamento** antes, durante ou depois do **lançamento**.	lançador.n: [frm_athletes_by_sport] [frm_athletes_by_position] desclassificar.v: [frm_referee_actions] zona de lançamento.n: [frm_sport_venues_subparts] lançamento.n: [frm_individual_moves] [frm_interactive_moves] [frm_sports_disciplines]	The **athlete** is **disqualified** if he/she leaves the **circle** before, during or after the **throw**.	athlete.n: [frm_athletes] disqualify.v: [frm_referee_actions] circle.n: [frm_sport_venues_subparts] throw.n: [frm_individual_moves] [frm_interactive_moves] [frm_sports_disciplines] [frm_winning_moves]	$\frac{5}{9} = 0{,}55$
3	O **árbitro** Mário Yamasaki **decidiu interromper** a **luta** achando que o **lutador** havia desmaiado.	árbitro.n: [frm_referees] decidir.v: [frm_referee_actions] interromper.v: [frm_referee_actions] luta.n: [frm_game] lutador.n: [frm_athletes_by_sport]	The **referee** Mario Yamasaki **decided** to **stop** the **fight** because he thought that the **fighter** had passed out.	referee.n: [frm_referees] decide.v: [frm_referee_actions] stop.v: [frm_referee_actions] fight.n: [frm_game] fighter.n: [frm_athletes_by_sport]	$\frac{5}{5} = 1$
4	Coloco o ponto em que o **levantador** executa o **levantamento**.	levantador.n: [frm_athletes_by_sport] [frm_athletes_by_position] levantamento.n: [frm_individual_moves] [frm_interactive_moves]	The sticking point at which the **setter** performs the **lift**.	setter.n: [frm_athletes_by_position] lift.n: [frm_individual_moves]	$\frac{2}{4} = 0{,}5$
5	O **ponta** é o **jogador** que menos tempo tem para pensar na armação de uma **jogada**.	ponta.n: [frm_athletes_by_position] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] jogada.n: [frm_technical_tactical_strategies]	The **winger** is the **player** with less time to think about setting up a **strike**.	winger.n: [frm_athletes_by_position] player.n: [frm_athletes] [frm_people_by_leisure_activity] strike.n: [frm_technical_tactical_strategies]	$\frac{4}{4} = 1$
6	O **ginásio** possui uma **quadra** que pode receber **jogos** de **futsal** e **handebol**.	ginásio.n: [frm_sport_venues] quadra.n: [frm_sport_venues] jogo.n: [frm_competition] [frm_game] futsal.n: [frm_sports] handebol.n: [frm_sports]	The **gym** has a **court** on which **futsal** and **handball games** can be played.	gym.n: [frm_sport_venues] court.n: [frm_sport_venues] futsal.n: [frm_sports] handball.n: [frm_sports] game.n: [frm_competition] [frm_game]	$\frac{6}{6} = 1$
7	O OG Kyle Long sofreu uma lesão na mão durante o primeiro **quarto** do **jogo** contra os Saints.	quarto.n: [frm_sport_temporal_subdivision] jogo.n: [frm_competition] [frm_game]	OG Kyle Long injured his hand during the first **quarter** against the Saints.	quarter.n: [frm_sport_temporal_subdivision]	$\frac{1}{3} = 0{,}33$
8	Além disso, a **rede** possui 1,55 cm sendo mais alta que a utilizada	rede.n: [frm_sport_venues_subparts] tênis.n: [frm_sports]	Also, the **net** is 1.55 cm higher than the one used in **tennis**	net.n: [frm_sport_venues_subparts] tennis.n: [frm_sports]	$\frac{6}{6} = 1$

	no **tênis** e menor que a **rede** da **quadra** de **vôlei**.	rede.n: [frm_sport_venues_subparts] quadra.n: [frm_sport_venues] vôlei.n: [frm_sports] quadra de vôlei.n: [frm_sport_venues]	and lower than the **net** used in the **volleyball court**.	net.n: [frm_sport_venues_subparts] volleyball.n: [frm_sports] court.n: [frm_sport_venues] volleyball court.n: [frm_sport_venues]	
9	A **competição** de **saltos** em **equipe** envolve um **saltador** e uma **saltadora**.	competição.n: [frm_competition] salto.n: [frm_individual_moves] [frm_sports] [frm_sports_disciplines] [frm_winning_moves] equipe.n: [frm_athletes] saltador.n: [frm_athletes_by_sport] saltadora.n: [frm_athletes_by_sport]	The **show jumping competition** involves a male and female **show jumper**.	show jumping.n: [frm_sports_disciplines] competition.n: [frm_competition] show jumper.n: [frm_athletes_by_sport] show jumper.n: [frm_athletes_by_sport]	$\frac{4}{8} = 0{,}5$
10	O **tênis**, um dos **esportes** mais tradicionais e praticados no mundo.	tênis.n: [frm_sports] esporte.n: [frm_sports]	**Tennis**, one of the most traditional **sports** played in the world.	tennis.n: [frm_sports] sport.n: [frm_sports] [frm_sports_scenario]	$\frac{2}{3} = 0{,}66$
11	A **bandeja** é quando o **jogador** faz a **cesta** bem próxima do **aro**.	bandeja.n: [frm_winning_moves] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] cesta.n: [frm_sport_venues_subparts] [frm_winning_moves] aro.n: [frm_sport_equipment] [frm_sport_venues_subparts]	The **layup** is when the **player** lays the **ball** off the **backboard** into the **hoop**.	layup.n: [frm_winning_moves] player.n: [frm_athletes] [frm_people_by_leisure_activity] ball.n: [frm_sport_equipment] backboard.n: [frm_sport_venues_subparts] hoop.n: [frm_sport_equipment] [frm_sport_venues_subparts]	$\frac{6}{8} = 0{,}75$
12	Durante uma **jogada**, um **jogador** pode dar até dois **toques** não consecutivos, de modo que a **equipe** só pode dar no total três **toques** na **bola**.	jogada.n: [frm_technical_tactical_strategies] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] toque.n: [frm_individual_moves] [frm_interactive_moves] [frm_winning_moves] equipe.n: [frm_athletes] toque.n: [frm_individual_moves] [frm_interactive_moves] [frm_winning_moves] bola.n: [frm_sport_equipment]	During **play**, a **player** can only **touch** the **ball** twice not consecutively, and the **team** can only **touch** the **ball** three times.	play.n: [frm_technical_tactical_strategies] player.n: [frm_athletes] [frm_people_by_leisure_activity] touch.v: [frm_individual_moves] [frm_winning_moves] ball.n: [frm_sport_equipment] team.n: [frm_athletes] touch.v: [frm_individual_moves] [frm_winning_moves] ball.n: [frm_sport_equipment]	$\frac{9}{12} = 0{,}75$
13	Mario Suárez **chuta** rente à **trave** do **gol** de Schwarzer.	chutar.v: [frm_individual_moves] [frm_interactive_moves] trave.n: [frm_sport_equipment] [frm_sport_venues_subparts] gol.n: [frm_sport_venues_subparts] [frm_winning_moves]	Mario Suárez just misses the Schwarzer's **goal post**.	goal.n: [frm_sport_venues_subparts] [frm_winning_moves] post.n: [frm_sport_venues_subparts]	$\frac{3}{6} = 0{,}5$
14	A **vela** é um **esporte olímpico** desde 1900.	vela.n: [frm_sports] [frm_sports_equipment] esporte.n: [frm_sports] olímpico.a: [frm_game]	14 - **Sailing** has been an **Olympic sport** since 1900.	sailing.n: [frm_sports] Olympic.a: [frm_game] sport.n: [frm_sports] [frm_sports_scenario]	$\frac{3}{5} = 0{,}6$
15	No último **lance** do **jogo**, o **zagueiro** Gustavo Gómez deu um	lance.n: [frm_technical_tactical_strategies]	15 - At the **end** of the **game**, **defender** Gustavo Gómez	end.n: [frm_sport_temporal_subdivision]	$\frac{7}{9} = 0{,}77$

	carrinho no **atacante** corinthiano Jô e fez o **pênalti**.	jogo.n: [frm_competition] [frm_game] zagueiro.n: [frm_athletes_by_position] carrinho.n: [frm_direct_infractions] atacante.n: [frm_athletes_by_position] pênalti.n: [frm_individual_moves] [frm_sanctions]	**slide tackled** the Corinthians **center forward** Jô and a **penalty kick** was given.	game.n: [frm_competition] [frm_game] defender.n: [frm_athletes_by_position] slide tackle.v: [frm_direct_infractions] center forward.n: [frm_athletes_by_position] penalty kick.n: [frm_individual_moves] [frm_sanctions]	
16	O **artilheiro** é o **jogador** Evandro, que **marcou** sete **gols**.	artilheiro.n: [frm_athletes_by_position] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] marcar.v: [frm_interactive_moves] [frm_referee_actions] [frm_winning_moves] gol.n: [frm_sport_venues_subp arts] [frm_winning_moves]	16 - Evandro was the **top scorer**, having **scored** seven **goals**.	top scorer.n: [frm_athletes_by_position] score.v: [frm_winning_moves] goal.n: [frm_sport_venues_subp arts] [frm_winning_moves]	$\frac{4}{8} = 0{,}5$
17	**Fuga** é uma **técnica** utilizada no **ciclismo de estrada**.	fuga.n: [frm_individual_moves] técnica.n: [frm_technical_tactical_strategies] ciclismo.n: [frm_sports] ciclismo de estrada.n: [frm_sports_disciplines]	17 - **Breakaway** is a **technique** used in **road cycling**.	breakaway.n: [frm_individual_moves] technique.n: [frm_technical_tactical_strategies] cycling.n: [frm_sports] road cycling.n: [frm_sports_disciplines]	$\frac{4}{4} = 1$
18	A **partida** se inicia sempre com um **saque**, **jogada** que, obrigatoriamente, alterna-se entre os **participantes** a cada **game**.	partida.n: [frm_game] saque.n: [frm_individual_moves] jogada.n: [frm_technical_tactical_strategies] participante.n: [frm_athletes] game.n: [frm_competition] [frm_game]	18 - The **game** always starts with a **serve**, which must be alternated between the **participants** at the beginning of a new **game**.	game.n: [frm_competition] [frm_game] serve.n: [frm_individual_moves] participant.n: [frm_athletes] game.n: [frm_competition] [frm_game]	$\frac{5}{7} = 0{,}71$
19	**Jogando** na **posição** 3, um **ala** é o **jogador** que mais se aproxima dos dois extremos das **posições** do **basquete**.	jogar.v: [frm_competition] [frm_technical_tactical_strategies] posição.n: [frm_athletes_by_position] ala.n: [frm_athletes_by_position] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] posição.n: [frm_athletes_by_position] basquete.n: [frm_sports]	19 - The **small forward**, number 3, is the **player** who comes closer to the two extremes of the **basketball positions**.	small forward.n: [frm_athletes_by_position] player.n: [frm_athletes] [frm_people_by_leisure_activity] basketball.n: [frm_sports] position.n: [frm_athletes_by_position]	$\frac{5}{8} = 0{,}62$
20	A **equipe jogava** bem e Juciely, com uma **china**, **jogada** muito utilizada pela **atleta**, fez 09-06.	equipe.n: [frm_athletes] jogar.v: [frm_competition] [frm_technical_tactical_strategies] china.n: [frm_individual_moves] jogada.n: [frm_technical_tactical_strategies]	20 - The **team played** well and Juciely, on a **slide**, a **play** used by the **athlete**, made it 09-06.	team.n: [frm_athletes] play.v: [frm_competition] [frm_technical_tactical_strategies] slide.n: [frm_individual_moves] play.n: [frm_technical_tactical_strategies]	$\frac{6}{7} = 0{,}85$

		atleta.n: [frm_athletes] [frm_people_by_vocation]		athlete.n: [frm_athletes]	
21	No **jogo de volta** da **decisão**, **torcedores** do River apedrejaram o ônibus dos **jogadores** do Boca Juniors, que não tiveram condições de entrar no **campo** do Estádio Monumental de Nuñez, em Buenos Aires.	jogo.n: [frm_competition] [frm_game] jogo de volta.n: [frm_game] decisão.n: [frm_referee_actions] torcedor.n: [frm_crowd] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] campo.n: [frm_sport_venues] [frm_sport_venues_subparts]	21 - On the decisive **second leg**, River **fans** stoned the bus of the Boca Juniors **players**, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	second leg.n: [frm_game] fan.n: [frm_crowd] player.n: [frm_athletes] [frm_people_by_leisure_activity]	$\frac{4}{9} = 0{,}44$
22	Com 18 anos, o **capitão** do **time** foi o **jogador** mais jovem da história do **clube** a **marcar** em um Atletiba.	capitão.n: [frm_athletes_by_position] [frm_people_by_vocation] time.n: [frm_athletes] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] clube.n: [frm_athletes] marcar.v: [frm_interactive_moves] [frm_referee_actions] [frm_winning_moves]	22 - At 18, the **captain** of the **team** was the youngest **player** in the history of the **club** to **score** in an Atletiba (Atlético x Curitiba).	captain.n: [frm_athletes_by_position] [frm_people_by_vocation] team.n: [frm_athletes] player.n: [frm_athletes] [frm_people_by_leisure_activity] club.n: [frm_sport_equipment] [frm_athletes] score.v: [frm_winning_moves]	$\frac{8}{9} = 0{,}88$
23	Então a **piscina** tem de ser dividida em dez **raias** para que só as oito internas, menos turbulentas, sejam usadas nas **provas**.	piscina.n: [frm_sport_venues_subparts] raia.n: [frm_sport_venues_subparts] prova.n: [frm_game]	23 - So the **swimming pool** has to be divided into ten **lanes** so that only the eight internal, less turbulent, are used in the **relays**.	swimming pool.n: [frm_sport_venues_subparts] lane.n: [frm_sport_venues_subparts] relay.n: [frm_game]	$\frac{3}{3} = 1$
24	O teoricamente dono da **posição** está novamente sem atuar, desta vez por lesão, enquanto o **reserva** é o **jogador** que mais vezes atuou nesta **temporada**.	posição.n: [frm_athletes_by_position] reserva.n: [frm_athletes_by_position] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] temporada.n: [frm_game]	24 - Theoretically, the owner of the **position** is unable to **play**, this time due to an injury, while the **reserve** is the **player** who has **played** the most this **season**.	position.n: [frm_athletes_by_position] play.v: [frm_competition] [frm_technical_tactical_strategies] reserve.n: [frm_athletes_by_position] player.n: [frm_athletes] [frm_people_by_leisure_activity] play.v: [frm_competition] [frm_technical_tactical_strategies] season.n: [frm_game]	$\frac{5}{9} = 0{,}55$
25	O **cruzamento** é uma **jogada** forte nossa, como é de todas as **equipes**.	cruzamento.n: [frm_interactive_moves] jogada.n: [frm_technical_tactical_strategies] equipe.n: [frm_athletes]	25 - The **crossing** is our strongest skill, like all the **teams**.	crossing.n: [frm_interactive_moves] team.n: [frm_athletes]	$\frac{2}{3} = 0{,}66$
26	No entanto, a mídia **esportiva** costuma referir à **jogada** apenas como **bicicleta**, pouco empregando o prefixo **chute** ou **pontapé**.	esportivo.a: [frm_sports] jogada.n: [frm_technical_tactical_strategies] bicicleta.n: [frm_individual_moves] [frm_sport_equipment] chute.n: [frm_individual_moves]	26 - However, the **sports** media usually refers to it as a **bicycle**, not using the prefix **kick** or **shot**.	sports.a: [frm_sports] bicycle.n: [frm_sport_equipment] kick.n: [frm_individual_moves] [frm_interactive_moves] shot.n: [frm_individual_moves] [frm_sport_equipment]	$\frac{6}{8} = 0{,}75$

		[frm_interactive_moves] pontapé.n: [frm_individual_moves] [frm_interactive_moves]			
27	O **estádio** possui uma **pista de atletismo** de **raias**, dois **telões** gigantes e uma rede wi-fi de última geração.	estádio.n: [frm_sport_venues] atletismo.n: [frm_sports] pista.n: [frm_sport_venues] [frm_sport_venues_subparts] pista de atletismo.n: [frm_sport_venues] raia.n: [frm_sport_venues_subparts] telão.n: [frm_sport_venues_subparts]	27 - The **stadium** has an **athletics track** of nine **lanes**, two giant **screens** and a state-of-the-art wi-fi network.	stadium.n: [frm_sport_venues] athletics.n: [frm_sports] track.n: [frm_sport_venues] athletics track.n: [frm_sport_venues] lane.n: [frm_sport_venues_subparts] screen.n: [frm_sport_venues_subparts]	$\frac{6}{7} = 0{,}85$
28	Quatro países irão **disputar** o **desafio** dos quatro **estilos** da **natação**, onde **borboleta** é o **nado** mais complexo de aprender.	disputar.v: [frm_competition] [frm_elimination_tournament] desafio.n: [frm_competition] estilo.n: [frm_technical_tactical_strategies] natação.n: [frm_sports] borboleta.n: [frm_individual_moves] [frm_sports_disciplines] nado.n: [frm_individual_moves]	28 - Four countries will **compete** in the **challenge** of four **swimming strokes**, where **butterfly** is the most complex **stroke** to learn.	compete.v: [frm_competition] [frm_elimination_tournament] challenge.n: [frm_competition] swimming.n: [frm_sports] stroke.n: [frm_athletes_by_position] [frm_individual_moves] [frm_interactive_moves] [frm_technical_tactical_strategies] butterfly.n: [frm_individual_moves] [frm_sports_disciplines] stroke.n: [frm_athletes_by_position] [frm_individual_moves] [frm_interactive_moves] [frm_technical_tactical_strategies]	$\frac{8}{14} = 0{,}57$
29	Segundo dados do Footstats, o **ponta** foi o **jogador** com mais **cruzamentos** certos e mais **passes** para **gol** na **equipe**.	ponta.n: [frm_athletes_by_position] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] cruzamento.n: [frm_interactive_moves] passe.n: [frm_interactive_moves] gol.n: [frm_sport_venues_subparts] [frm_winning_moves] equipe.n: [frm_athletes]	29 - According to data from Footstats, the **winger** was the **player** with the best **crosses** and more **assists** in the **team**.	winger.n: [frm_athletes_by_position] player.n: [frm_athletes] [frm_people_by_leisure_activity] cross.n: [frm_interactive_moves] [frm_winning_moves] assist.n: [frm_winning_moves] team.n: [frm_athletes]	$\frac{6}{9} = 0{,}66$
30	Totalmente diferente dos outros **estilos**, o **nado peito** exige muita coordenação e **técnica** do praticante.	estilo.n: [frm_technical_tactical_strategies] nado.n: [frm_individual_moves] peito.n: [frm_individual_moves] [frm_sports_disciplines] técnica.n: [frm_technical_tactical_strategies]	30 - Totally different from other **strokes**, **breaststroke** requires a lot of coordination and **technique** from the **swimmer**.	stroke.n: [frm_athletes_by_position] [frm_individual_moves] [frm_interactive_moves] [frm_technical_tactical_strategies] breaststroke.n: [frm_individual_moves] [frm_sports_disciplines] technique.n: [frm_technical_tactical_strategies] swimmer.n: [frm_athletes_by_sport]	$\frac{5}{8} = 0{,}62$

31	Pedrinho dá um **chapéu**, mas não dá sequência na **jogada**, aos doze minutos do primeiro **tempo**.	chapéu.n: [frm_interactive_moves] jogada.n: [frm_technical_tactical_strategies] tempo.n: [frm_sport_temporal_subdivision]	31 - Pedrinho **lobs** the **player**, but does not follow through, at twelve minutes of the first **half**.	lob.v: [frm_interactive_moves] player.n: [frm_athletes] [frm_people_by_leisure_activity] half.n: [frm_sport_temporal_subdivision] [frm_winning_moves]	$\frac{2}{6} = 0{,}33$
32	A tradicional comemoração com um **peixinho** na **quadra** está viva na memória.	peixinho.n: [frm_individual_moves] quadra.n: [frm_sport_venues]	32 - The traditional celebration with a **dive** on the **court** is alive in the memory.	dive.n: [frm_individual_moves] court.n: [frm_sport_venues]	$\frac{2}{2} = 1$
33	A posição e o alinhamento da **gaiola** no **campo** de **competição** é, portanto, crítico para o seu uso seguro.	gaiola.n: [frm_sport_venues_subparts] campo.n: [frm_sport_venues] [frm_sport_venues_subparts] competição.n: [frm_competition]	33 - The position and alignment of the **cage** in the **competition field** is, therefore, critical for their safe use.	cage.n: [frm_sport_venues_subparts] competition.n: [frm_competition] field.n: [frm_sport_venues]	$\frac{3}{4} = 0{,}75$
34	O **nado** de **costas** causa uma boa sensação após a execução de **séries** intensas de **crawl**, ou **livre**, e **borboleta**.	nado.n: [frm_individual_moves] costas.n: [frm_individual_moves] [frm_sports_disciplines] série.n: [frm_sport_temporal_subdivision] crawl.n: [frm_individual_moves] [frm_sports_disciplines] livre.n: [frm_individual_moves] borboleta.n: [frm_individual_moves] [frm_sports_disciplines]	34 - The **backstroke** causes a good feeling after the execution of intense **series** of **crawl**, or **freestyle**, and **butterfly**.	backstroke.n: [frm_individual_moves] [frm_sports_disciplines] series.n: [frm_sport_temporal_subdivision] crawl.n: [frm_individual_moves] [frm_sports_disciplines] freestyle.n: [frm_individual_moves] [frm_sports_disciplines] butterfly.n: [frm_individual_moves] [frm_sports_disciplines]	$\frac{9}{9} = 1$
35	Muitos diziam que este era um **salto** criado pela escola queniana de **atletismo**, mas me parece que é uma variante do **salto tesoura**.	salto.n: [frm_individual_moves] [frm_sports] [frm_sports_disciplines] [frm_winning_moves] atletismo.n: [frm_sports] salto tesoura.n: [frm_individual_moves]	35 - Many said that this **jump** was created by the Kenyan school of **athletics**, but it seems to me that it is a variant of the **scissors jump**.	jump.n: [frm_individual_moves] [frm_winning_moves] athletics.n: [frm_sports] scissors jump.n: [frm_individual_moves]	$\frac{4}{6} = 0{,}66$
36	Muitos acham que o **gancho** é o **golpe** onde o **lutador** lança sua mão de baixo para cima.	gancho.n: [frm_interactive_moves] golpe.n: [frm_interactive_moves] lutador.n: [frm_athletes]	36 - Many people think that the **hook** is the **punch** where the **fighter punches** from the bottom upwards.	hook.n: [frm_interactive_moves] punch.n: [frm_interactive_moves] fighter.n: [frm_athletes] punch.v: [frm_interactive_moves]	$\frac{3}{4} = 0{,}75$
37	Companheiro de Messi no Barcelona, o **meia** é o **jogador** brasileiro com o maior valor a atuar na Copa América.	meia.n: [frm_athletes_by_position] jogador.n: [frm_athletes] [frm_people_by_leisure_activity]	37 - Messi's team mate in Barcelona, the **midfielder** is the highest valued Brazilian **player** to **play** in the Copa América.	midfielder.n: [frm_athletes_by_position] player.n: [frm_athletes] [frm_people_by_leisure_activity] play.v: [frm_competition] [frm_technical_tactical_strategies]	$\frac{3}{5} = 0{,}6$
38	Cinco **séries** de **golpes** combinados de **suple**, **bombeiro** e **estabilização** no solo com ênfase na precisão de movimento.	golpe.n: [frm_interactive_moves] suple.n: [frm_winning_moves] bombeiro.n: [frm_winning_moves] estabilização.n: [frm_winning_moves]	38 - Five combined series of **blows** of **suplex**, **fireman carry's slam** and **pin** down with emphasis on precision of movement.	blow.n: [frm_interactive_moves] suplex.n: [frm_winning_moves] fireman carry's slam.n: [frm_winning_moves] pin.n: [frm_winning_moves]	$\frac{4}{4} = 1$

39	No **rugby**, o **abertura** é o **jogador** mais habilidoso do **time**.	rugby.n: [frm_sports] abertura.n: [frm_athletes_by_position] [frm_medal_cerimony] [frm_winning_moves] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] time.n: [frm_athletes]	39 - In **rugby**, the **fly-half** is the most skilled **player** in the **team**.	rugby.n: [frm_sports] fly-half.n: [frm_athletes_by_position] player.n: [frm_athletes] [frm_people_by_leisure_activity] team.n: [frm_athletes]	$\frac{5}{7} = 0{,}71$
40	O **servidor** é o **jogador** que coloca a **bola** em **jogo** para o primeiro **ponto**.	servidor.n: [frm_athletes_by_position] [frm_people_by_vocation] jogador.n: [frm_athletes] [frm_people_by_leisure_activity] bola.n: [frm_sport_equipment] jogo.n: [frm_competition] [frm_game] ponto.n: [frm_winning_moves]	40 - The **server** is the **player** who puts the **ball** into **play** for the first **point**.	server.n: [frm_athletes_by_position] [frm_people_by_vocation] player.n: [frm_athletes] [frm_people_by_leisure_activity] ball.n: [frm_sport_equipment] play.n: [frm_technical_tactical_strategies] point.n: [frm_winning_moves]	$\frac{6}{9} = 0{,}66$
41	O sonho de Tristan Garcia, de 14 anos e **fã** de **basquetebol** era **arremessar** uma **bola** na **cesta** da **quadra** da escola.	fã.n: [frm_crowd] basquetebol.n: [frm_sports] arremessar.v: [frm_individual_moves] [frm_interactive_moves] bola.n: [frm_sport_equipment] cesta.n: [frm_sport_venues_subparts] [frm_winning_moves] quadra.n: [frm_sport_venues]	41 - The dream of Tristan Garcia, aged 14 and a **fan** of **basketball** was **throwing** a **ball** into the **basket** of the school **court**.	fan.n: [frm_crowd] basketball.n: [frm_sports] throw.v: [frm_finals_play] [frm_individual_moves] [frm_interactive_moves] ball.n: [frm_sport_equipment] basket.n: [frm_sport_venues_subparts] [frm_winning_moves] court.n: [frm_sport_venues]	$\frac{8}{9} = 0{,}88$
42	Um **arco** é um **equipamento** individual e pessoal.	arco.n: [frm_sport_equipment] [frm_winning_moves] equipamento.n: [frm_sport_equipment]	42 - A **bow** is for individual and personal use.	bow.n: [frm_athletes_by_position] [frm_sport_equipment]	$\frac{1}{4} = 0{,}25$
43	Invista em uma boa **luva**, a **luva** é o **equipamento** de segurança maior do **boxe**, ela pode evitar que você se machuque gravemente.	luva.n: [frm_clothing] luva.n: [frm_clothing] equipamento.n: [frm_sport_equipment] boxe.n: [frm_sports]	43 - Invest in a good **glove**, the **glove** is the greatest safety **equipment** in **boxing**, it can avoid you getting hurt badly.	glove.n: [frm_clothing] glove.n: [frm_clothing] equipmento.n: [frm_sport_equipment] boxing.n: [frm_sports]	$\frac{4}{4} = 1$
44	Todos os **ginastas** que **disputam** a **prova saltam** sobre um **aparelho** ligeiramente inclinado chamado **mesa**.	ginasta.n: [frm_athletes_by_sport] disputar.v: [frm_competition] [frm_elimination_tournament] prova.n: [frm_game] saltar.v: [frm_individual_moves] [frm_winning_moves] aparelho.n: [frm_sport_equipment] mesa.n: [frm_sport_equipment] [frm_sport_venues_subparts]	44 - All **gymnasts** who **compete jump** on an **apparatus** slightly inclined called a **vault**.	gymnast.n: [frm_athletes_by_sport] compete.v: [frm_competition] [frm_elimination_tournament] jump.v: [frm_individual_moves] [frm_winning_moves] apparatus.n: [frm_sport_equipment] vault.n: [frm_individual_moves] [frm_sports_disciplines] [frm_sport_equipment]	$\frac{7}{11} = 0{,}63$
45	Se a **ginástica rítmica** é a ovelha negra da família **ginástica**, em seguida, a **corda** é a ovelha negra dos **aparelhos**.	ginástica.n: [frm_sports] ginástica rítmica.n: [frm_sports_disciplines] ginástica.n: [frm_sports] corda.n: [frm_sport_equipment]	45 - If **rhythmic gymnastics** is the black sheep of the **gymnastics** family, then the **rope** is the	gymnastics.n: [frm_sports] rhythmic gymnastics.n: [frm_sports_disciplines] gymnastics.n: [frm_sports]	$\frac{5}{5} = 1$

		aparelho.n: [frm_sport_equipment]	black sheep of the **equipment**.	rope.n: [frm_sport_equipment] equipment.n: [frm_sport_equipment]	
46	A **fita** é considerada o **aparelho** mais plástico e característico da **ginástica**.	fita.n: [frm_sport_equipment] aparelho.n: [frm_sport_equipment] ginástica.n: [frm_sports]	46 - The **ribbon** is considered the most plastic component and characteristic of **gymnastics**.	ribbon.n: [frm_sport_equipment] gymnastics.n: [frm_sports]	$\frac{2}{3} = 0{,}66$
47	Ela **venceu** as **disputas** nos **aparelhos**: **fita** e **maças**.	vencer.v: [frm_finals_play] disputa.n: [frm_competition] aparelho.n: [frm_sport_equipment] fita.n: [frm_sport_equipment] maça.n: [frm_sport_equipment]	47 - She **won** the **competition** on the **apparatus**: **ribbon** and **clubs**.	win.v: [frm_finals_play] competition.n: [frm_competition] apparatus.n: [frm_sport_equipment] ribbon.n: [frm_sport_equipment] club.n: [frm_athletes] [frm_sport_equipment]	$\frac{5}{6} = 0{,}83$
48	A **vara** para **salto** é um **equipamento** muito avançado.	vara.n: [frm_sport_equipment] salto.n: [frm_individual_moves] [frm_sports] [frm_sports_disciplines] [frm_winning_moves] equipamento.n: [frm_sport_equipment]	48 - The **vaulting pole** is a very advanced **equipment**.	vault.v: [frm_individual_moves] pole.n: [frm_sport_equipment] equipment.n: [frm_sport_equipment]	$\frac{3}{6} = 0{,}5$
49	O **atacante** ainda protagonizou um **lance** de brilho ao aplicar um **lençol** em um **adversário**, **jogada** que chamou atenção de outros **boleiros** na web.	atacante.n: [frm_athletes_by_position] lance.n: [frm_technical_tactical_strategies] lençol.n: [frm_interactive_moves] adversário.n: [frm_athletes] jogada.n: [frm_technical_tactical_strategies] boleiro.n: [frm_athletes_by_sport]	49 - The **center forward** also showed brilliance on **lobbing** the **ball** over his **opponent**, which drew the attention of other **footballers** on the web.	center forward.n: [frm_athletes_by_position] lob.v: [frm_interactive_moves] ball.n: [frm_sport_equipment] opponent.n: [frm_athletes] footballer.n: [frm_athletes_by_sport]	$\frac{4}{7} = 0{,}57$
50	A **prancha** é o **equipamento** mais importante para **pegar ondas** grandes.	prancha.n: [frm_sport_equipment] equipamento.n: [frm_sport_equipment] pegar.v: [frm_individual_moves] [frm_winning_moves] onda.n: [frm_weather] [frm_winning_moves]	50 - The **board** is the most important **equipment** to **catch** big **waves**.	board.n: [frm_sport_equipment] [frm_sport_venues_subparts] equipment.n: [frm_sport_equipment] catch.v: [frm_individual_moves] [frm_winning_moves] wave.n: [frm_weather] [frm_winning_moves]	$\frac{6}{7} = 0{,}85$
Total de frames correspondentes evocados tanto pelas sentenças-fonte quanto pelas sentenças-alvo em inglês traduzidas (Padrão de Referência). A porcentagem de correspondência semântica é calculada tomando-se a soma do número de frames correspondentes evocados pelas ULs da sentença-fonte e sentença-alvo. Faz-se uma divisão pelo número total de frames diferentes dos esportes por evocados pelas ULs de ambas das sentenças somados.					$\frac{36{,}2}{50} = 72{,}4\ \%$

Fonte: Compilado pelo autor (2020).

APÊNDICE C – AVALIAÇÃO DE TM BLEU – S-BASE

Tabela 11 – Avaliação de TM BLEU aplicada nas traduções do Sistema de TM Estado da Arte – Google Tradutor (S-Base)

Tipo de Tradução	Sentença Traduzida	BLEU Score
Ref.	A runner does not try to run a marathon in the first days of training.	84.92
S-Base	A runner does not try to run a marathon in the first few days of training.	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	39.03
S-Base	The pitcher is disqualified if he leaves the launch zone before, during or after the launch.	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	63.70
S-Base	Referee Mario Yamasaki decided to stop the fight thinking that the fighter had passed out.	
Ref.	The sticking point at which the setter performs the lift.	40.01
S-Base	I place the point at which the lifter performs the lift.	
Ref.	The winger is the player with less time to think about setting up a strike.	51.61
S-Base	The forward is the player who has less time to think about setting up a move.	
Ref.	The gym has a court on which futsal and handball games can be played.	52.03
S-Base	The gym has a court that can host futsal and handball games.	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	42.84
S-Base	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	54.63
S-Base	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net on the volleyball court.	
Ref.	The show jumping competition involves a male and female show jumper.	29.63
S-Base	The team jumping competition involves a jumper and a jumper.	
Ref.	Tennis, one of the most traditional sports played in the world.	68.65
S-Base	Tennis, one of the most traditional and practiced sports in the world.	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	21.34
S-Base	The tray is when the player makes the basket very close to the ring.	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	19.56
S-Base	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	
Ref.	Mario Suárez just misses the Schwarzer's goal post.	25.45
S-Base	Mario Suárez kicks close to the goal post of Schwarzer.	
Ref.	Sailing has been an Olympic sport since 1900.	100.0
S-Base	Sailing has been an Olympic sport since 1900.	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	28.15
S-Base	In the last game of the game, defender Gustavo Gómez gave the striker Corinthiano Jô a trolley and made the penalty.	
Ref.	Evandro was the top scorer, having scored seven goals.	27.88
S-Base	The top scorer is the player Evandro, who scored seven goals.	
Ref.	Breakaway is a technique used in road cycling.	74.19
S-Base	Escape is a technique used in road cycling.	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	45.38
S-Base	The game always starts with a serve, a move that must alternate between the participants in each game.	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	38.68
S-Base	Playing in position 3, a winger is the player who comes closest to the two ends of the basketball positions.	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	52.79
S-Base	The team played well and Juciely, with a china, played very often by the athlete, made 09-06.	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	43.63
S-Base	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	
Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	44.10
S-Base	At 18, the team captain was the youngest player in the club's history to score in an atletiba.	

Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	62.23
S-Base	Then the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	66.49
S-Base	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	
Ref.	The crossing is our strongest skill, like all the teams.	11.30
S-Base	The cross is a strong move for us, as it is for all teams.	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	44.40
S-Base	However, the sports media usually refer to the play only as a bicycle, with little use of the prefix kick or kick.	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	75.45
S-Base	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	48.50
S-Base	Four countries will contest the challenge of the four styles of swimming, where butterfly is the most complex swim to learn.	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	42.15
S-Base	According to data from the Footstats, the forward was the player with the most correct crosses and the most passes for goals in the team.	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	60.27
S-Base	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	
Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	39.02
S-Base	Pedrinho gives a hat, but does not give sequence in the play, to the twelve minutes of the first half.	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	67.05
S-Base	The traditional celebration with a goldfish on the court is alive in memory.	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	66.35
S-Base	The position and alignment of the cage in the competition field is therefore critical to its safe use.	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	76.57
S-Base	The backstroke causes a good sensation after the execution of intense series of crawl, or free, and butterfly.	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	66.18
S-Base	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	40.05
S-Base	Many think that the hook is the blow where the fighter throws his hand from the bottom up.	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	46.72
S-Base	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the Copa America.	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	29.50
S-Base	Five series of combined strokes of suple, fireman and ground stabilization with an emphasis on precision of movement.	
Ref.	In rugby, the fly-half is the most skilled player in the team.	54.08
S-Base	In rugby, the aperture is the most skilled player on the team.	
Ref.	The server is the player who puts the ball into play for the first point.	83.94
S-Base	The server is the player who puts the ball in play for the first point.	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	51.78
S-Base	The dream of Tristan Garcia, age 14 and a basketball fan, was to throw a ball into the basket on the school court.	
Ref.	A bow is for individual and personal use.	48.95
S-Base	A bow is individual and personal equipment.	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	65.22
S-Base	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	

Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	20.59
S-Base	All gymnasts competing in the competition jump on a slightly inclined device called a table.	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	92.44
S-Base	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	68.92
S-Base	The ribbon is considered the most plastic and characteristic apparatus of gymnastics.	
Ref.	She won the competition on the apparatus: ribbon and clubs.	40.01
S-Base	She won the disputes on the devices: ribbon and apples.	
Ref.	The vaulting pole is a very advanced equipment.	38.31
S-Base	The jumping pole is very advanced equipment.	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	25.31
S-Base	The striker also made a brilliant move by applying a sheet to an opponent, a move that drew the attention of other players on the web.	
Ref.	The board is the most important equipment to catch big waves.	69.97
S-Base	The board is the most important equipment for catching big waves.	
BLEU Total Médio S-Base = 53.13		

Fonte: Compilado pelo autor (2020).

APÊNDICE D – AVALIAÇÃO DE TM BLEU – S-PRÉ

Tabela 12 – Avaliação de TM BLEU aplicada nas traduções do Sistema de TM Semanticamente Enriquecido com Injeção Terminológica no Pré-processamento (S-Pré)

Tipo de Tradução	Sentença Traduzida	BLEU Score
Ref.	A runner does not try to run a marathon in the first days of training.	66.72
S-Pré	A racer does not try to run a marathon during the first days of training.	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	39.03
S-Pré	The thrower is disqualified if he leaves the throwing zone before, during or after the throwing.	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	41.98
S-Pré	Judge Mário Yamasaki decided to stop the combat thinking the fighter had passed out.	
Ref.	The sticking point at which the setter performs the lift.	25.40
S-Pré	I place the point at which the lifter performs the facelift.	
Ref.	The winger is the player with less time to think about setting up a strike.	23.18
S-Pré	The wing is the player that has less time to think in the setup of a play.	
Ref.	The gym has a court on which futsal and handball games can be played.	44.81
S-Pré	The gym has a court that can host a game of futsal and handball.	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	42.84
S-Pré	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	23.10
S-Pré	In addition, the net has 1.55 cm being higher than that used in tennis and smaller than the net of a volleyball court.	
Ref.	The show jumping competition involves a male and female show jumper.	14.56
S-Pré	The team jump dispute involves a vaulter and a vaulter.	
Ref.	Tennis, one of the most traditional sports played in the world.	68.65
S-Pré	Tennis, one of the most traditional and practiced sports in the world.	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	47.82
S-Pré	The layup is when the player makes the basket very close to the hoop.	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	22.66
S-Pré	During a play, a player can give up to two non-consecutive touches, so the team can only give a total of three touches on a ball.	
Ref.	Mario Suárez just misses the Schwarzer’s goal post.	22.22
S-Pré	Mario Suárez kicks close to the post of Schwarzer's goal.	
Ref.	Sailing has been an Olympic sport since 1900.	100.0
S-Pré	Sailing has been an Olympic sport since 1900.	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	16.08
S-Pré	In the last move of the game, center back Gustavo Gómez tackled the forward Corinthian Jô and made the penalty.	
Ref.	Evandro was the top scorer, having scored seven goals.	24.02
S-Pré	The leading scorer is player Evandro, who scored seven goals.	
Ref.	Breakaway is a technique used in road cycling.	100.0
S-Pré	Breakaway is a technique used in road cycling.	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	35.72
S-Pré	The game always starts with a service, play that, obligatorily, alternates between the participants in each game.	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	24.76
S-Pré	Playing in position 3, a winger is the player that most closely matches the two extremes of the position of the basketball.	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	50.04
S-Pré	The team played well and Juciely, with a slide, play widely used by the sportsman, made 09-06.	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	27.31
S-Pré	In the return game of the decision, supporter of the River stoned the bus of the players of the Boca Juniors, who had no condition to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	

Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	71.11
S-Pré	At 18, the captain of the club was the youngest player in the history of the club to score in an Atletiba.	
Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	85.12
S-Pré	Then the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the event.	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	66.49
S-Pré	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	
Ref.	The crossing is our strongest skill, like all the teams.	11.30
S-Pré	The cross is a strong play for us, as it is for all teams.	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	38.20
S-Pré	However, the sports media usually refer to play only as a scissor kick, with little use of the prefix kick or kicking.	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	68.86
S-Pré	The stadium has a nine lane athletics track, two giant screens and a state-of-the-art wi-fi net.	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	50.64
S-Pré	Four countries will compete for the challenge of the four stroke of a swimming, where fly is the most complex stroke to learn.	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	37.01
S-Pré	According to data from the Footstats, the wing was the player with more certain crosses and more pass for goal en a team.	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	60.27
S-Pré	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	
Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	29.91
S-Pré	Pedrinho gives a lob, but does not give sequence in a play, to the twelve minutes of the first time.	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	60.25
S-Pré	The traditional celebration with a dive en a court is alive in memory.	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	39.72
S-Pré	The position and alignment of a cage in the dispute field is therefore critical to its safe use.	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	23.06
S-Pré	The back stroke causes a good sensation after performing intense crawl, or free, and butterfly routines.	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	78.63
S-Pré	Many said that this was a jump created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissor jump.	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	40.05
S-Pré	Many think that the hook is the stroke where the fighter throws his hand from the bottom up.	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	43.94
S-Pré	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the America Cup.	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	24.46
S-Pré	Five combined stroke routines of suplex, fireman's carry and hold on the floor with an emphasis on movement accuracy.	
Ref.	In rugby, the fly-half is the most skilled player in the team.	89.15
S-Pré	In rugby, the fly-half is the most skilled player in the club.	
Ref.	The server is the player who puts the ball into play for the first point.	60.31
S-Pré	The server is the player that puts the ball in game for the first point.	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	51.89
S-Pré	The dream of Tristan Garcia, aged 14 and a basketball fan, was to throw a ball into a basket from a school court.	
Ref.	A bow is for individual and personal use.	25.21
S-Pré	A hoop is an individual and personal apparatus.	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	51.13

S-Pré	Invest in a good glove, the glove is the biggest security guard apparatus of boxing, it can prevent you from getting seriously hurt.	
Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	16.67
S-Pré	All gymnast competing in the event jump on a slightly inclined apparatus called a vault table.	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	82.07
S-Pré	If the rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	60.57
S-Pré	The ribbon is considered the most plastic and characteristic apparatus of a gymnastics.	
Ref.	She won the competition on the apparatus: ribbon and clubs.	62.98
S-Pré	She won the contest on the apparatus: ribbon and club.	
Ref.	The vaulting pole is a very advanced equipment.	32.64
S-Pré	The pole for jump is a very advanced apparatus.	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	25.31
S-Pré	The forward also featured a brilliant move to apply a lob to an adversary, a play that drew the attention of other players on the web.	
Ref.	The board is the most important equipment to catch big waves.	56.22
S-Pré	The board is the most important apparatus for catching large waves.	
BLEU Total Médio S-Pré = 48.12		

Fonte: Compilado pelo autor (2020).

APÊNDICE E – AVALIAÇÃO DE TM BLEU – S-PÓS

Tabela 13 – Avaliação de TM BLEU aplicada nas traduções do Sistema de TM Semanticamente Enriquecido com Injeção Terminológica na Pós-edição (S-Pós)

Tipo de Tradução	Sentença Traduzida	BLEU Score
Ref.	A runner does not try to run a marathon in the first days of training.	84.92
S-Pós	A runner does not try to run a marathon in the first few days of training.	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	39.03
S-Pós	The thrower is disqualified if he leaves the launch zone before, during or after the release.	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	39.14
S-Pós	Ref Mário Yamasaki decided to suspend the fight thinking that the fighter had passed out.	
Ref.	The sticking point at which the setter performs the lift.	65.91
S-Pós	I place the point at which the setter performs the lift.	
Ref.	The winger is the player with less time to think about setting up a strike.	68.59
S-Pós	The winger is the player who has less time to think about setting up a play.	
Ref.	The gym has a court on which futsal and handball games can be played.	52.03
S-Pós	The gym has a court that can receive futsal and handball games.	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	42.84
S-Pós	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	54.63
S-Pós	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net of the volleyball court.	
Ref.	The show jumping competition involves a male and female show jumper.	29.63
S-Pós	The team jumping competition involves a jumper and a jumper.	
Ref.	Tennis, one of the most traditional sports played in the world.	68.65
S-Pós	Tennis, one of the most traditional and practiced sports in the world.	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	42.28
S-Pós	The layup is when the player makes the basket very close to the ring.	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	19.56
S-Pós	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	
Ref.	Mario Suárez just misses the Schwarzer’s goal post.	39.83
S-Pós	Mario Suárez shoots close to Schwarzer's goal post.	
Ref.	Sailing has been an Olympic sport since 1900.	100.0
S-Pós	Sailing has been an Olympic sport since 1900.	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	29.55
S-Pós	In the last game of the game, defender Gustavo Gómez gave a tackle to Corinthians striker Jô and made the penalty.	
Ref.	Evandro was the top scorer, having scored seven goals.	27.88
S-Pós	The top scorer is the player Evandro, who scored seven goals.	
Ref.	Breakaway is a technique used in road cycling.	100.0
S-Pós	Breakaway is a technique used in road cycling.	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	18.96
S-Pós	The match always begins with a serve, a move that must alternate between the participants in each match.	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	38.68
S-Pós	Playing in position 3, a winger is the player who comes closest to the two ends of the basketball positions.	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	56.43
S-Pós	The team played well and Juciely, with a slide, played by the athlete, made 09-06.	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	41.58
S-Pós	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental de Nuñez Stadium in Buenos Aires.	
Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	44.10

S-Pós	At 18, the team captain was the youngest player in the club's history to score in an Atletiba.	
Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	67.83
S-Pós	So the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	53.26
S-Pós	The theoretically owner of the position is again out of action, stays time due to injury, while the reserve is the player who has played the most stays season.	
Ref.	The crossing is our strongest skill, like all the teams.	8.51
S-Pós	Crossing is a strong move for us, as it is for all teams.	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	39.25
S-Pós	However, the sports media usually refer to the play only as a bicycle kick, with little use of the prefix kick or kick.	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	75.45
S-Pós	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	66.94
S-Pós	Four countries will compete in the challenge of the four swimming styles, where butterfly is the most complex swim to learn.	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	51.49
S-Pós	According to data from Footstats, the forward was the player with the most correct crosses and the most passing for goals in the team.	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	72.00
S-Pós	Totally different from other strokes, breast stroke requires a lot of coordination and technique from the practitioner.	
Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	39.02
S-Pós	Pedrinho gives a lob, but does not give sequence in the play, to the twelve minutes of the first half.	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	85.21
S-Pós	The traditional celebration with a dive on the court is alive in memory.	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	52.83
S-Pós	The position and alignment of the cage on the competition field is therefore critical to its safe use.	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	54.97
S-Pós	The backstroke causes a good sensation after the execution of intense routine of freestyle, or free, and butterfly.	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	66.18
S-Pós	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	40.05
S-Pós	Many think that the hook is the blow where the fighter throws his hand from the bottom up.	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	44.73
S-Pós	Messi's companion at Barcelona, the midfielder is the Brazilian player with the highest value to play in the America Cup.	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	37.74
S-Pós	Five routine of combined strokes of suplex, fireman's carry and floor hold with an emphasis on precision of movement.	
Ref.	In rugby, the fly-half is the most skilled player in the team.	54.08
S-Pós	In rugby, the opener is the most skilled player on the team.	
Ref.	The server is the player who puts the ball into play for the first point.	77.78
S-Pós	The server is the player who puts the ball in game for the first point.	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	45.05
S-Pós	The dream of Tristan Garcia, a 14-year-old basketball fan, was to throw a ball into the basket on the school court.	
Ref.	A bow is for individual and personal use.	48.95
S-Pós	A bow is individual and personal equipment.	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	65.22

S-Pós	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	
Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	20.59
S-Pós	All gymnasts competing in the competition jump on a slightly inclined equipment called a table.	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	92.44
S-Pós	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	68.92
S-Pós	The ribbon is considered the most plastic and characteristic equipment of gymnastics.	
Ref.	She won the competition on the apparatus: ribbon and clubs.	45.93
S-Pós	She won the disputes on ribbon and clubs.	
Ref.	The vaulting pole is a very advanced equipment.	70.16
S-Pós	The jumping pole is a very advanced equipment.	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	25.31
S-Pós	The striker also made a brilliant play by applying a lob to an opponent, a play that drew the attention of other players on the web.	
Ref.	The board is the most important equipment to catch big waves.	69.97
S-Pós	The board is the most important equipment for catching big waves.	
BLEU Total Médio S-Pós = 53.66		

Fonte: Compilado pelo autor (2020).

APÊNDICE F – AVALIAÇÃO DE TM TER – S-BASE

Tabela 14 – Avaliação de TM TER aplicada nas traduções do Sistema de TM Estado da Arte – Google Tradutor (S-Base)

Tipo de Tradução	Sentença Traduzida	TER Score
Ref.	A runner does not try to run a marathon in the first days of training.	6.667
S-Base	A runner does not try to run a marathon in the first few days of training.	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	33.33
S-Base	The pitcher is disqualified if he leaves the launch zone before, during or after the launch.	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	22.22
S-Base	Referee Mario Yamasaki decided to stop the fight thinking that the fighter had passed out.	
Ref.	The sticking point at which the setter performs the lift.	40.0
S-Base	I place the point at which the lifter performs the lift.	
Ref.	The winger is the player with less time to think about setting up a strike.	26.66
S-Base	The forward is the player who has less time to think about setting up a move.	
Ref.	The gym has a court on which futsal and handball games can be played.	42.85
S-Base	The gym has a court that can host futsal and handball games.	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	46.15
S-Base	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	21.73
S-Base	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net on the volleyball court.	
Ref.	The show jumping competition involves a male and female show jumper.	36.36
S-Base	The team jumping competition involves a jumper and a jumper.	
Ref.	Tennis, one of the most traditional sports played in the world.	27.27
S-Base	Tennis, one of the most traditional and practiced sports in the world.	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	53.33
S-Base	The tray is when the player makes the basket very close to the ring.	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	77.27
S-Base	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	
Ref.	Mario Suárez just misses the Schwarzer's goal post.	87.50
S-Base	Mario Suárez kicks close to the goal post of Schwarzer.	
Ref.	Sailing has been an Olympic sport since 1900.	0
S-Base	Sailing has been an Olympic sport since 1900.	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	59.09
S-Base	In the last game of the game, defender Gustavo Gómez gave the striker Corinthiano Jô a trolley and made the penalty.	
Ref.	Evandro was the top scorer, having scored seven goals.	77.77
S-Base	The top scorer is the player Evandro, who scored seven goals.	
Ref.	Breakaway is a technique used in road cycling.	12.50
S-Base	Escape is a technique used in road cycling.	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	47.61
S-Base	The game always starts with a serve, a move that must alternate between the participants in each game.	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	42.10
S-Base	Playing in position 3, a winger is the player who comes closest to the two ends of the basketball positions.	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	33.33
S-Base	The team played well and Juciely, with a china, played very often by the athlete, made 09-06.	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	50.0
S-Base	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	
Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	44.0
S-Base	At 18, the team captain was the youngest player in the club's history to score in an atletiba.	

Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	25.0
S-Base	Then the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	24.13
S-Base	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	
Ref.	The crossing is our strongest skill, like all the teams.	110.0
S-Base	The cross is a strong move for us, as it is for all teams.	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	50.0
S-Base	However, the sports media usually refer to the play only as a bicycle, with little use of the prefix kick or kick.	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	35.2
S-Base	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	35.0
S-Base	Four countries will contest the challenge of the four styles of swimming, where butterfly is the most complex swim to learn.	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	45.0
S-Base	According to data from the Footstats, the forward was the player with the most correct crosses and the most passes for goals in the team.	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	25.0
S-Base	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	
Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	62.50
S-Base	Pedrinho gives a hat, but does not give sequence in the play, to the twelve minutes of the first half.	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	14.28
S-Base	The traditional celebration with a goldfish on the court is alive in memory.	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	22.22
S-Base	The position and alignment of the cage in the competition field is therefore critical to its safe use.	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	15.78
S-Base	The backstroke causes a good sensation after the execution of intense series of crawl, or free, and butterfly.	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	22.22
S-Base	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	35.29
S-Base	Many think that the hook is the blow where the fighter throws his hand from the bottom up.	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	36.84
S-Base	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the Copa America.	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	42.10
S-Base	Five series of combined strokes of suple, fireman and ground stabilization with an emphasis on precision of movement.	
Ref.	In rugby, the fly-half is the most skilled player in the team.	16.66
S-Base	In rugby, the aperture is the most skilled player on the team.	
Ref.	The server is the player who puts the ball into play for the first point.	6.66
S-Base	The server is the player who puts the ball in play for the first point.	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	30.43
S-Base	The dream of Tristan Garcia, age 14 and a basketball fan, was to throw a ball into the basket on the school court.	
Ref.	A bow is for individual and personal use.	25.0
S-Base	A bow is individual and personal equipment.	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	23.81
S-Base	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	

Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	61.53
S-Base	All gymnasts competing in the competition jump on a slightly inclined device called a table.	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	4.76
S-Base	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	16.66
S-Base	The ribbon is considered the most plastic and characteristic apparatus of gymnastics.	
Ref.	She won the competition on the apparatus: ribbon and clubs.	30.0
S-Base	She won the disputes on the devices: ribbon and apples.	
Ref.	The vaulting pole is a very advanced equipment.	25.0
S-Base	The jumping pole is very advanced equipment.	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	69.56
S-Base	The striker also made a brilliant move by applying a sheet to an opponent, a move that drew the attention of other players on the web.	
Ref.	The board is the most important equipment to catch big waves.	18.18
S-Base	The board is the most important equipment for catching big waves.	
TER Total Médio S-Base = 36.23		

Fonte: Compilado pelo autor (2020).

APÊNDICE G – AVALIAÇÃO DE TM TER – S-PRÉ

Tabela 15 – Avaliação de TM TER aplicada nas traduções do Sistema de TM Semanticamente Enriquecido com Injeção Terminológica no Pré-processamento (S-Pré)

Tipo de Tradução	Sentença Traduzida	TER Score
Ref.	A runner does not try to run a marathon in the first days of training.	13.33
S-Pré	A racer does not try to run a marathon during the first days of training.	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	33.33
S-Pré	The thrower is disqualified if he leaves the throwing zone before, during or after the throwing.	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	44.44
S-Pré	Judge Mário Yamasaki decided to stop the combat thinking the fighter had passed out.	
Ref.	The sticking point at which the setter performs the lift.	50.0
S-Pré	I place the point at which the lifter performs the facelift.	
Ref.	The winger is the player with less time to think about setting up a strike.	53.33
S-Pré	The wing is the player that has less time to think in the setup of a play.	
Ref.	The gym has a court on which futsal and handball games can be played.	57.14
S-Pré	The gym has a court that can host a game of futsal and handball.	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	46.15
S-Pré	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	39.13
S-Pré	In addition, the net has 1.55 cm being higher than that used in tennis and smaller than the net of a volleyball court.	
Ref.	The show jumping competition involves a male and female show jumper.	63.63
S-Pré	The team jump dispute involves a vaulter and a vaulter.	
Ref.	Tennis, one of the most traditional sports played in the world.	27.27
S-Pré	Tennis, one of the most traditional and practiced sports in the world.	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	40.0
S-Pré	The layup is when the player makes the basket very close to the hoop.	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	77.27
S-Pré	During a play, a player can give up to two non-consecutive touches, so the team can only give a total of three touches on a ball.	
Ref.	Mario Suárez just misses the Schwarzer’s goal post.	87.50
S-Pré	Mario Suárez kicks close to the post of Schwarzer's goal.	
Ref.	Sailing has been an Olympic sport since 1900.	0
S-Pré	Sailing has been an Olympic sport since 1900.	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	59.09
S-Pré	In the last move of the game, center back Gustavo Gómez tackled the forward Corinthian Jô and made the penalty.	
Ref.	Evandro was the top scorer, having scored seven goals.	77.77
S-Pré	The leading scorer is player Evandro, who scored seven goals.	
Ref.	Breakaway is a technique used in road cycling.	0
S-Pré	Breakaway is a technique used in road cycling.	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	52.38
S-Pré	The game always starts with a service, play that, obligatorily, alternates between the participants in each game.	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	68.42
S-Pré	Playing in position 3, a winger is the player that most closely matches the two extremes of the position of the basketball.	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	27.77
S-Pré	The team played well and Juciely, with a slide, play widely used by the sportsman, made 09-06.	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	85.71
S-Pré	In the return game of the decision, supporter of the River stoned the bus of the players of the Boca Juniors, who had no condition to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	

Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	20.00
S-Pré	At 18, the captain of the club was the youngest player in the history of the club to score in an Atletiba.	
Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	8.33
S-Pré	Then the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the event.	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	24.13
S-Pré	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	
Ref.	The crossing is our strongest skill, like all the teams.	110.0
S-Pré	The cross is a strong play for us, as it is for all teams.	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	55.55
S-Pré	However, the sports media usually refer to play only as a scissor kick, with little use of the prefix kick or kicking.	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	41.17
S-Pré	The stadium has a nine lane athletics track, two giant screens and a state-of-the-art wi-fi net.	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	35.0
S-Pré	Four countries will compete for the challenge of the four stroke of a swimming, where fly is the most complex stroke to learn.	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	45.0
S-Pré	According to data from the Footstats, the wing was the player with more certain crosses and more pass for goal en a team.	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	25.0
S-Pré	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	
Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	68.75
S-Pré	Pedrinho gives a lob, but does not give sequence in a play, to the twelve minutes of the first time.	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	21.42
S-Pré	The traditional celebration with a dive en a court is alive in memory.	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	33.33
S-Pré	The position and alignment of a cage in the dispute field is therefore critical to its safe use.	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	63.15
S-Pré	The back stroke causes a good sensation after performing intense crawl, or free, and butterfly routines.	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	11.11
S-Pré	Many said that this was a jump created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissor jump.	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	35.29
S-Pré	Many think that the hook is the stroke where the fighter throws his hand from the bottom up.	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	42.10
S-Pré	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the America Cup.	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	73.68
S-Pré	Five combined stroke routines of suplex, fireman's carry and hold on the floor with an emphasis on movement accuracy.	
Ref.	In rugby, the fly-half is the most skilled player in the team.	8.33
S-Pré	In rugby, the fly-half is the most skilled player in the club.	
Ref.	The server is the player who puts the ball into play for the first point.	20.0
S-Pré	The server is the player that puts the ball in game for the first point.	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	34.78
S-Pré	The dream of Tristan Garcia, aged 14 and a basketball fan, was to throw a ball into a basket from a school court.	
Ref.	A bow is for individual and personal use.	37.50
S-Pré	A hoop is an individual and personal apparatus.	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	42.85

S-Pré	Invest in a good glove, the glove is the biggest security guard apparatus of boxing, it can prevent you from getting seriously hurt.	
Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	69.23
S-Pré	All gymnast competing in the event jump on a slightly inclined apparatus called a vault table.	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	9.52
S-Pré	If the rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	25.0
S-Pré	The ribbon is considered the most plastic and characteristic apparatus of a gymnastics.	
Ref.	She won the competition on the apparatus: ribbon and clubs.	20.0
S-Pré	She won the contest on the apparatus: ribbon and club.	
Ref.	The vaulting pole is a very advanced equipment.	50.0
S-Pré	The pole for jump is a very advanced apparatus.	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	69.56
S-Pré	The forward also featured a brilliant move to apply a lob to an adversary, a play that drew the attention of other players on the web.	
Ref.	The board is the most important equipment to catch big waves.	36.36
S-Pré	The board is the most important apparatus for catching large waves.	
TER Total Médio S-Pré = 42.63		

Fonte: Compilado pelo autor (2020).

APÊNDICE H – AVALIAÇÃO DE TM TER – S-PÓS

Tabela 16 – Avaliação de TM TER aplicada nas traduções do Sistema de TM Semanticamente Enriquecido com Injeção Terminológica na Pós-edição (S-Pós)

Tipo de Tradução	Sentença Traduzida	TER Score
Ref.	A runner does not try to run a marathon in the first days of training.	6.66
S-Pós	A runner does not try to run a marathon in the first few days of training.	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	33.33
S-Pós	The thrower is disqualified if he leaves the launch zone before, during or after the release.	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	38.88
S-Pós	Ref Mário Yamasaki decided to suspend the fight thinking that the fighter had passed out.	
Ref.	The sticking point at which the setter performs the lift.	30.0
S-Pós	I place the point at which the setter performs the lift.	
Ref.	The winger is the player with less time to think about setting up a strike.	20.0
S-Pós	The winger is the player who has less time to think about setting up a play.	
Ref.	The gym has a court on which futsal and handball games can be played.	42.85
S-Pós	The gym has a court that can receive futsal and handball games.	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	46.15
S-Pós	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	21.73
S-Pós	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net of the volleyball court.	
Ref.	The show jumping competition involves a male and female show jumper.	36.36
S-Pós	The team jumping competition involves a jumper and a jumper.	
Ref.	Tennis, one of the most traditional sports played in the world.	27.27
S-Pós	Tennis, one of the most traditional and practiced sports in the world.	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	46.66
S-Pós	The layup is when the player makes the basket very close to the ring.	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	77.27
S-Pós	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	
Ref.	Mario Suárez just misses the Schwarzer’s goal post.	50.0
S-Pós	Mario Suárez shoots close to Schwarzer's goal post.	
Ref.	Sailing has been an Olympic sport since 1900.	0
S-Pós	Sailing has been an Olympic sport since 1900.	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	59.09
S-Pós	In the last game of the game, defender Gustavo Gómez gave a tackle to Corinthians striker Jô and made the penalty.	
Ref.	Evandro was the top scorer, having scored seven goals.	77.77
S-Pós	The top scorer is the player Evandro, who scored seven goals.	
Ref.	Breakaway is a technique used in road cycling.	0
S-Pós	Breakaway is a technique used in road cycling.	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	61.90
S-Pós	The match always begins with a serve, a move that must alternate between the participants in each match.	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	42.10
S-Pós	Playing in position 3, a winger is the player who comes closest to the two ends of the basketball positions.	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	27.77
S-Pós	The team played well and Juciely, with a slide, played by the athlete, made 09-06.	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	50.0
S-Pós	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental de Nuñez Stadium in Buenos Aires.	
Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	44.0

S-Pós	At 18, the team captain was the youngest player in the club's history to score in an Atletiba.	
Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	20.8
S-Pós	So the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	31.03
S-Pós	The theoretically owner of the position is again out of action, stays time due to injury, while the reserve is the player who has played the most stays season.	
Ref.	The crossing is our strongest skill, like all the teams.	110.0
S-Pós	Crossing is a strong move for us, as it is for all teams.	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	61.11
S-Pós	However, the sports media usually refer to the play only as a bicycle kick, with little use of the prefix kick or kick.	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	35.29
S-Pós	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	15.0
S-Pós	Four countries will compete in the challenge of the four swimming styles, where butterfly is the most complex swim to learn.	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	40.0
S-Pós	According to data from Footstats, the forward was the player with the most correct crosses and the most passing for goals in the team.	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	18.75
S-Pós	Totally different from other strokes, breast stroke requires a lot of coordination and technique from the practitioner.	
Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	62.50
S-Pós	Pedrinho gives a lob, but does not give sequence in the play, to the twelve minutes of the first half.	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	7.14
S-Pós	The traditional celebration with a dive on the court is alive in memory.	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	27.77
S-Pós	The position and alignment of the cage on the competition field is therefore critical to its safe use.	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	26.31
S-Pós	The backstroke causes a good sensation after the execution of intense routine of freestyle, or free, and butterfly.	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	22.22
S-Pós	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	35.29
S-Pós	Many think that the hook is the blow where the fighter throws his hand from the bottom up.	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	47.36
S-Pós	Messi's companion at Barcelona, the midfielder is the Brazilian player with the highest value to play in the America Cup.	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	47.36
S-Pós	Five routine of combined strokes of suplex, fireman's carry and floor hold with an emphasis on precision of movement.	
Ref.	In rugby, the fly-half is the most skilled player in the team.	16.66
S-Pós	In rugby, the opener is the most skilled player on the team.	
Ref.	The server is the player who puts the ball into play for the first point.	13.33
S-Pós	The server is the player who puts the ball in game for the first point.	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	39.13
S-Pós	The dream of Tristan Garcia, a 14-year-old basketball fan, was to throw a ball into the basket on the school court.	
Ref.	A bow is for individual and personal use.	25.0
S-Pós	A bow is individual and personal equipment.	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	23.81

S-Pós	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	
Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	61.53
S-Pós	All gymnasts competing in the competition jump on a slightly inclined equipment called a table.	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	4.76
S-Pós	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	16.66
S-Pós	The ribbon is considered the most plastic and characteristic equipment of gymnastics.	
Ref.	She won the competition on the apparatus: ribbon and clubs.	30.0
S-Pós	She won the disputes on ribbon and clubs.	
Ref.	The vaulting pole is a very advanced equipment.	12.50
S-Pós	The jumping pole is a very advanced equipment.	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	69.56
S-Pós	The striker also made a brilliant play by applying a lob to an opponent, a play that drew the attention of other players on the web.	
Ref.	The board is the most important equipment to catch big waves.	18.18
S-Pós	The board is the most important equipment for catching big waves.	
TER Total Médio S-Pós = 36.47		

Fonte: Compilado pelo autor (2020).

APÊNDICE I – AVALIAÇÃO DE TM HTER – S-BASE

Tabela 17 – Avaliação de TM HTER – Edições Humanas Feitas nas traduções do Sistema de TM Estado da Arte – Google Tradutor (S-Base)

Referência / Editores	Tradução de Referência (Quantidade de Palavras) / Edições das traduções S-Base	Quantidade de Palavras e Edições		HTER Scores	HTER Score Médio
Ref.	A runner does not try to run a marathon in the first days of training.	Palavras	15		
S-Base	A runner does not try to run a marathon in the first few days of training.	Edições			
Ed.1	A runner does not try to run a marathon in the first ~~few~~ days of training.	Ins.: 0, Ap.: 1, Sub.: 0, MPos.: 0	1	$\frac{1}{15} = 0,06$	$\frac{0,32}{3} = 0,10$
Ed.2	A runner does not try to run a marathon in the first ~~few~~ days of training.	Ins.: 0, Ap.: 1, Sub.: 0, MPos.: 0	1	$\frac{1}{15} = 0,06$	
Ed.3	A runner does not try to run a marathon in the first ~~few days of~~ training days.	Ins.: 0, Ap.: 2, Sub.: 0, MPos.: 1	3	$\frac{3}{15} = 0,2$	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	Palavras	15		
S-Base	The pitcher is disqualified if he leaves the launch zone before, during or after the launch.	Edições			
Ed.1	The pitcher is disqualified if he/she leaves the launch zone before, during or after the launch.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{15} = 0,06$	$\frac{0,45}{3} = 0,15$
Ed.2	The ~~pitcher~~ athlete is disqualified if he/she leaves the ~~launch~~ start zone before, during or after the ~~launch~~ start.	Ins.: 0, Ap.: 0, Sub.: 4, MPos.: 0	4	$\frac{4}{15} = 0,26$	
Ed.3	The pitcher is disqualified if he leaves the ~~launch zone~~ launching circle before, during or after the launch.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{15} = 0,13$	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	Palavras	18		
S-Base	Referee Mario Yamasaki decided to stop the fight thinking that the fighter had passed out.	Edições			
Ed.1	Referee Mario Yamasaki decided to stop the fight thinking that the fighter had passed out.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	$\frac{0,22}{3} = 0,07$
Ed.2	Referee Mario Yamasaki decided to stop the fight thinking that the fighter had passed out.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	
Ed.3	Referee Mario Yamasaki decided to stop the fight ~~thinking~~ because he thought that the fighter had passed out.	Ins.: 2, Ap.: 0 Sub.: 1, MPos.: 1	4	$\frac{4}{18} = 0,22$	
Ref.	The sticking point at which the setter performs the lift.	Palavras	10		
S-Base	I place the point at which the lifter performs the lift.	Edições			
Ed.1	I place the point at which the lifter performs the lift.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	$\frac{0,8}{3} = 0,26$
Ed.2	~~I place the~~ It's the sticking point at which the lifter performs the lift.	Ins.: 3, Ap.: 2, Sub.: 0, MPos.: 0	5	$\frac{5}{10} = 0,5$	
Ed.3	~~I place the~~ The sticking point at which the lifter performs the lift.	Ins.: 1, Ap.: 2, Sub.: 0, MPos.: 0	3	$\frac{3}{10} = 0,3$	
Ref.	The winger is the player with less time to think about setting up a strike.	Palavras	15		
S-Base	The forward is the player who has less time to think about setting up a move.	Edições			

Ed.1	The ~~forward~~ wingback is the player who has less time to think about setting up a move.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{15} = 0,06$	$\frac{0,26}{3} = 0,08$
Ed.2	The forward is the player who has less time to think about setting up a move.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ed.3	The ~~forward~~ winger is the player who has less time to think about setting up a striking shot ~~move~~.	Ins.: 1, Ap.: 0, Sub.: 2, MPos.: 0	3	$\frac{3}{15} = 0,2$	
Ref.	The gym has a court on which futsal and handball games can be played.	Palavras	14		
S-Base	The gym has a court that can host futsal and handball games.	Edições			
Ed.1	The gym has a court that can host futsal and handball games.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{14} = 0$	$\frac{0,07}{3} = 0,02$
Ed.2	The gym has a court that can host futsal and handball games.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{14} = 0$	
Ed.3	The gym has a court that can host futsal and handball ~~games~~ matches.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{14} = 0,07$	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	Palavras	13		
S-Base	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	Edições			
Ed.1	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{13} = 0$	$\frac{0,07}{3} = 0,02$
Ed.2	OG Kyle ~~l~~Long suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0,07$	
Ed.3	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{13} = 0$	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	Palavras	23		
S-Base	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net on the volleyball court.	Edições			
Ed.1	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net on the volleyball court.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	$\frac{0,17}{3} = 0,05$
Ed.2	In addition, the net is 1.55 cm taller than the one used in tennis and ~~smaller~~ lower than the net ion the volleyball court.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0,04$	
Ed.3	In addition, the net is 1.55 cm ~~taller~~ higher than the one used in tennis and ~~smaller~~ lower than the net ~~on~~ in the volleyball court.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{23} = 0,13$	
Ref.	The show jumping competition involves a male and female show jumper.	Palavras	11		
S-Base	The team jumping competition involves a jumper and a jumper.	Edições			
Ed.1	The team jumping competition involves ~~a~~ man and ~~jumper~~ woman jumpers ~~and a jumper~~.	Ins.: 2, Ap.: 2, Sub.: 1, MPos.: 1	6	$\frac{6}{11} = 0,54$	$\frac{1,35}{3} = 0,45$
Ed.2	The team of show jumping competition involves a male jumper and a female jumper.	Ins.: 4, Ap.: 0, Sub.: 0, MPos.: 0	4	$\frac{4}{11} = 0,36$	
Ed.3	The ~~team~~ jumping competition show involves a male ~~jumper~~ and a female jumper.	Ins.: 3, Ap.: 2, Sub.: 0, MPos.: 0	5	$\frac{5}{11} = 0,45$	
Ref.	Tennis, one of the most traditional sports played in the world.	Palavras	11		
S-Base	Tennis, one of the most traditional and practiced sports in the world.	Edições			

Ed.1	Tennis, one of the most traditional and practiced sports in the world.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	$\frac{0,45}{3} = 0,15$
Ed.2	Tennis, one of the most traditional and ~~practiced~~ played sports in the world.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{11} = 0,09$	
Ed.3	Tennis, ~~one of~~ among the most traditional ~~and practiced~~ sports in the world.	Ins.: 0, Ap.: 3, Sub.: 1, MPos.: 0	4	$\frac{4}{11} = 0,36$	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	Palavras	15		
S-Base	The tray is when the player makes the basket very close to the ring.	Edições			
Ed.1	The layup ~~tray~~ is when the player ~~makes the basket~~ scores very close to the ring.	Ins.: 0, Ap.: 2, Sub.: 2, MPos.: 0	4	$\frac{4}{15} = 0,26$	$\frac{1,59}{3} = 0,53$
Ed.2	The layup ~~tray~~ is when the player ~~makes~~ hits the basket off the backboard into ~~very close to~~ the ring.	Ins.: 3, Ap.: 2, Sub.: 3, MPos.: 0	8	$\frac{8}{15} = 0,53$	
Ed.3	The ~~tray~~ layup is when the player bounces the ball off the backboard into the basket ~~makes the basket very close to the ring~~.	Ins.: 3, Ap.: 3, Sub.: 3, MPos.: 3	12	$\frac{12}{15} = 0,8$	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	Palavras	22		
S-Base	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	Edições			
Ed.1	During a rally, a player can make up to two non-consecutive hits, so the team can only make ~~a total of~~ ~~three~~ three hits on the ball.	Ins.: 0, Ap.: 3, Sub.: 0, MPos.: 0	3	$\frac{3}{22} = 0,13$	$\frac{0,26}{3} = 0,08$
Ed.2	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{22} = 0$	
Ed.3	During the play ~~a rally~~, a player can make up to two non-consecutive hits, ~~so~~ and the team can only make a total of three hits on the ball.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{22} = 0,13$	
Ref.	Mario Suárez just misses the Schwarzer's goal post.	Palavras	8		
S-Base	Mario Suárez kicks close to the goal post of Schwarzer.	Edições			
Ed.1	Mario Suárez kicks close to the Schwarzer's goal ~~goal post of Schwarzer~~.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 2	4	$\frac{4}{8} = 0,5$	$\frac{0,75}{3} = 0,25$
Ed.2	Mario Suárez kicks close to the goal post of Schwarzer.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	Mario Suárez kicks ~~close to~~ out of the goal post of Schwarzer.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{8} = 0,25$	
Ref.	Sailing has been an Olympic sport since 1900.	Palavras	8		
S-Base	Sailing has been an Olympic sport since 1900.	Edições			
Ed.1	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0		$\frac{0}{8} = 0$	$\frac{0}{3} = 0$
Ed.2	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0		$\frac{0}{8} = 0$	
Ed.3	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0		$\frac{0}{8} = 0$	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	Palavras	22		
S-Base	In the last game of the game, defender Gustavo Gómez gave the striker Corinthiano Jô a trolley and made the penalty.	Edições			
Ed.1	In the last game of the game, defender Gustavo Gómez ~~gave~~ hit the Corinthians	Ins.: 1, Ap.: 2, Sub.: 3, MPos.: 2	8	$\frac{8}{22} = 0,36$	$\frac{1,35}{3} = 0,45$

	striker ~~Corinthiano~~ Jô ~~a trolley~~ and a penalty was made ~~made the penalty~~.				
Ed.2	In the last ~~game~~ moment of the game, defender Gustavo Gómez slide tackled ~~gave~~ the Corinthians striker ~~Corinthiano~~ Jô ~~a trolley~~ and made the penalty.	Ins.: 1, Ap.: 2, Sub.: 3, MPos.: 1	7	$\frac{7}{22} = 0{,}31$	
Ed.3	In the last ~~game~~ minutes of the game, defender Gustavo Gómez slide tackled ~~gave the striker Corinthiano~~ Corinthian's player Jô and a ~~a trolley and made the~~ penalty kick was given.	Ins.: 4, Ap.: 3, Sub.: 6, MPos.: 2	15	$\frac{15}{22} = 0{,}68$	
Ref.	Evandro was the top scorer, having scored seven goals.	Palavras	9		
S-Base	The top scorer is the player Evandro, who scored seven goals.	Edições			
Ed.1	The top scorer is the player Evandro, who scored seven goals.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{9} = 0$	$\frac{0{,}22}{3} = 0{,}07$
Ed.2	The top scorer is the player Evandro, who has scored seven goals.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{9} = 0{,}11$	
Ed.3	The top scorer ~~is~~ was the player Evandro, who scored seven goals.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{9} = 0{,}11$	
Ref.	Breakaway is a technique used in road cycling.	Palavras	8		
S-Base	Escape is a technique used in road cycling.	Edições			
Ed.1	~~Escape~~ Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0{,}12$	$\frac{0{,}37}{3} = 012$
Ed.2	~~Escape~~ Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0{,}12$	
Ed.3	~~Escape~~ Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0{,}12$	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	Palavras	21		
S-Base	The game always starts with a serve, a move that must alternate between the participants in each game.	Edições			
Ed.1	The game always starts with a serve, a move that must alternate between the participants in each game.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	$\frac{0{,}33}{3} = 0{,}11$
Ed.2	The game always starts with a serve, a move that must alternate between the participants in the beginning of each game.	Ins.: 3, Ap.: 0, Sub.: 0, MPos.: 0	3	$\frac{3}{21} = 0{,}14$	
Ed.3	The game always starts with a serve, a move that must alternate between the participants in the beginning of each new game.	Ins.: 4, Ap.: 0, Sub.: 0, MPos.: 0	4	$\frac{4}{21} = 0{,}19$	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	Palavras	19		
S-Base	Playing in position 3, a winger is the player who comes closest to the two ends of the basketball positions.	Edições			
Ed.1	Playing in position 3, a winger is the player who comes closest to the two extremes ~~ends~~ of the basketball positions.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{19} = 0{,}05$	$\frac{0{,}67}{3} = 0{,}22$
Ed.2	Playing in position 3, a small forward ~~winger~~ is the player who comes closer~~st~~ to the two ends of the basketball positions.	Ins.: 1, Ap.: 0, Sub.: 2, MPos.: 0	3	$\frac{3}{19} = 0{,}15$	
Ed.3	~~Playing in position~~ The number 3 is~~,~~ a ~~winger~~ small forward ~~is the~~ player who comes closer~~st~~ closer to the two ends of the basketball positions.	Ins.: 1, Ap.: 3, Sub.: 3, MPos.: 2	9	$\frac{9}{19} = 0{,}47$	

Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	Palavras	18		
S-Base	The team played well and Juciely, with a china, played very often by the athlete, made 09-06.	Edições			
Ed.1	The team played well and Juciely, with a slide ~~china~~, a play ~~played very often~~ used very often by the athlete, made 09-06.	Ins.: 2, Ap.: 0, Sub.: 2, MPos.: 2	6	$\frac{6}{18} = 0,33$	$\frac{0,77}{3} = 0,25$
Ed.2	The team played well and Juciely, with a ~~china~~ slide, which is played very often by the athlete, made it 09-06.	Ins.: 3, Ap.: 0 Sub.: 1, MPos.: 0	4	$\frac{4}{18} = 0,22$	
Ed.3	The team played well and Juciely, with a ~~china~~ slide, commonly played ~~very often~~ by the athlete, made 09-06.	Ins.: 0, Ap.: 1, Sub.: 2, MPos.: 1	4	$\frac{4}{18} = 0,22$	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	Palavras	28		
S-Base	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	Edições			
Ed.1	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental de ~~Stadium of~~ Nuñez Stadium, in Buenos Aires.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 1	2	$\frac{2}{28} = 0,07$	$\frac{0,42}{3} = 0,14$
Ed.2	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{28} = 0$	
Ed.3	In the decisive return game ~~of the decision~~, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Estádio Monumental de Nuñez ~~Monumental Stadium of Nuñez~~, in Buenos Aires.	Ins.: 1, Ap.: 3, Sub.: 4, MPos.: 2	10	$\frac{10}{28} = 0,35$	
Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	Palavras	25		
S-Base	At 18, the team captain was the youngest player in the club's history to score in an atletiba.	Edições			
Ed.1	At 18, the team captain was the youngest player in the club's history to score in an "~~atletiba~~ Atletiba".	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{25} = 0,04$	$\frac{0,36}{3} = 0,12$
Ed.2	At 18, the team captain was the youngest player in the club's history to score in an ~~atletiba~~ Atletiba.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{25} = 0,04$	
Ed.3	At the age of 18, the team captain was the youngest player in the club's history to score in an ~~atletiba~~ Atletiba (Atlético x Curitiba).	Ins.: 6, Ap.: 0, Sub.: 1, MPos.: 0	7	$\frac{7}{25} = 0,28$	
Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	Palavras	24		
S-Base	Then the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	Edições			

Ed.1	Then, the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{24} = 0{,}04$	$\frac{0{,}2}{3} = 0{,}06$
Ed.2	Then the pool has to be divided into ten lanes, so that only the eight inner ~~indoor,~~ less turbulent ones~~,~~ are used in the events.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 1	3	$\frac{3}{24} = 0{,}12$	
Ed.3	Then the pool has to be divided into ten lanes so that only the eight ones indoor, less turbulent ~~ones~~, are used in the events.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 1	1	$\frac{1}{24} = 0{,}04$	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	Palavras	29		
S-Base	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	Edições			
Ed.1	The ~~theoretically~~ owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season, theoretically.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 1	2	$\frac{2}{29} = 0{,}06$	$\frac{0{,}19}{3} = 0{,}06$
Ed.2	The theoretical ~~theoretically~~ owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{29} = 0{,}03$	
Ed.3	~~The theoretically~~ In theory, the owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 1	3	$\frac{3}{29} = 0{,}10$	
Ref.	The crossing is our strongest skill, like all the teams.	Palavras	10		
S-Base	The cross is a strong move for us, as it is for all teams.	Edições			
Ed.1	The crossing ~~cross~~ is a strong move for us, as it is for all teams.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{10} = 0{,}1$	$\frac{1}{3} = 0{,}33$
Ed.2	The crossing ~~cross~~ is ~~a~~ our strong move ~~for us~~, as it is for all the teams.	Ins.: 1, Ap.: 2, Sub.: 2, MPos.: 0	5	$\frac{5}{10} = 0{,}5$	
Ed.3	The crossing ~~cross~~ is ~~a~~ our strong move ~~for us~~, as it is for all teams.	Ins.: 0, Ap.: 2, Sub.: 2, MPos.: 0	4	$\frac{4}{10} = 0{,}4$	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	Palavras	18		
S-Base	However, the sports media usually refer to the play only as a bicycle, with little use of the prefix kick or kick.	Edições			
Ed.1	However, the sports media usually refer to the play only as a bicycle, ~~with~~ without ~~little~~ the use of the prefix kick or ~~kick~~ shoot.	Ins.: 1, Ap.: 1, Sub.: 2, MPos.: 0	4	$\frac{4}{18} = 0{,}22$	$\frac{0{,}38}{3} = 0{,}12$
Ed.2	However, the sports media usually refer to the play only as a bicycle, with little use of the prefix kick or ~~kick~~ shot.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{18} = 0{,}05$	
Ed.3	However, the sports media usually ~~refer~~ refers to the play only as a bicycle, with little use of the prefix kick or ~~kick~~ shot.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{18} = 0{,}11$	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	Palavras	17		
S-Base	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Edições			

Ed.1	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Ins.: 0, Ap.: 0, Sub.: 0, Pos.: 0	0	$\frac{0}{17} = 0$	$\frac{0}{3} = 0$
Ed.2	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Ins.: 0, Ap.: 0, Sub.: 0, Pos.: 0	0	$\frac{0}{17} = 0$	
Ed.3	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Ins.: 0, Ap.: 0, Sub.: 0, Pos.: 0	0	$\frac{0}{17} = 0$	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	Palavras	20		
S-Base	Four countries will contest the challenge of the four styles of swimming, where butterfly is the most complex swim to learn.	Edições			
Ed.1	Four countries will ~~contest~~ accept the challenge of the four styles of swimming, where butterfly is the most complex stroke ~~swim~~ to learn.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{20} = 0{,}1$	$\frac{0{,}2}{3} = 0{,}06$
Ed.2	Four countries will contest the challenge of the four styles of swimming, where butterfly is the most complex ~~swim~~ stroke to learn.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{20} = 0{,}05$	
Ed.3	Four countries will contest the challenge of the four styles of swimming, where butterfly is the most complex ~~swim~~ style to learn.	Ins.: 0, Ap.: 0 Sub.: 1, MPos.: 0	1	$\frac{1}{20} = 0{,}05$	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	Palavras	20		
S-Base	According to data from the Footstats, the forward was the player with the most correct crosses and the most passes for goals in the team.	Edições			
Ed.1	According to data from the Footstats, the forward was the player with the most correct crosses and the most passes for goals in the team.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{20} = 0$	$\frac{0{,}55}{3} = 0{,}18$
Ed.2	According to data from the Footstats, the forward was the player with the most correct crosses and the most passes to ~~for~~ goals in the team.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{20} = 0{,}05$	
Ed.3	According to data from the Footstats, the forward was the player with ~~the~~ more assists and had the most correct crosses ~~and the most passes for goals~~ in the team.	Ins.: 1, Ap.: 5, Sub.: 0, MPos.: 4	10	$\frac{10}{20} = 0{,}5$	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	Palavras	16		
S-Base	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	Edições			
Ed.1	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{16} = 0$	$\frac{0{,}18}{3} = 0{,}06$
Ed.2	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the ~~practitioner~~ swimmer.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{16} = 0{,}06$	
Ed.3	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the ~~practitioner~~ person swimming.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{16} = 0{,}12$	

Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	Palavras	16		
S-Base	Pedrinho gives a hat, but does not give sequence in the play, to the twelve minutes of the first half.	Edições			
Ed.1	Pedrinho ~~gives a hat~~ lobs the opponent, but does not give sequence in the play, to the twelve minutes of the first half.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{16} = 0{,}18$	$\frac{1{,}3}{3} = 0{,}43$
Ed.2	Pedrinho ~~gives a hat~~ lobbed the player, but did ~~does~~ not give sequence in the play, ~~to the~~ at twelve minutes of the first half.	Ins.: 0, Ap.: 1, Sub.: 5, MPos.: 0	6	$\frac{6}{16} = 0{,}37$	
Ed.3	Pedrinho dinks the ball over the opponent's head ~~gives a hat~~, but ~~does not~~ cannot give sequence in the play, ~~to the~~ at twelve minutes of the first half.	Ins.: 5, Ap.: 2, Sub.: 5, MPos.: 0	12	$\frac{12}{16} = 0{,}75$	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	Palavras	14		
S-Base	The traditional celebration with a goldfish on the court is alive in memory.	Edições			
Ed.1	The traditional celebration with a dive ~~goldfish~~ ~~on the court~~ is alive in our memory.	Ins.: 1, Ap.: 3, Sub.: 1, MPos.: 0	5	$\frac{5}{14} = 0{,}35$	$\frac{0{,}84}{3} = 0{,}28$
Ed.2	The traditional celebration with a ~~goldfish~~ dive on the court is alive in memory.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{14} = 0{,}07$	
Ed.3	The traditional celebration ~~with a goldfish~~ where they dive on the court is alive in memory.	Ins.: 3, Ap.: 3, Sub.: 0, MPos.: 0	6	$\frac{6}{14} = 0{,}42$	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	Palavras	18		
S-Base	The traditional celebration with a goldfish on the court is alive in memory.	Edições			
Ed.1	The position and alignment of the cage in the competition field is ~~therefore~~ critical to its safe use.	Ins.: 0, Ap.: 1, Sub.: 0, MPos.: 0	1	$\frac{1}{18} = 0{,}05$	$\frac{0{,}32}{3} = 0{,}1$
Ed.2	The position and alignment of the cage in the competition field is therefore critical to its safe use.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	
Ed.3	Therefore, the position and alignment of the cage in the competition field is ~~therefore~~ critical ~~to its~~ for a safe use.	Ins.: 2, Ap.: 1, Sub.: 1, MPos.: 1	5	$\frac{5}{18} = 0{,}27$	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	Palavras	19		
S-Base	The backstroke causes a good sensation after the execution of intense series of crawl, or free, and butterfly.	Edições			
Ed.1	The backstroke causes a good sensation after the execution of intense series of crawl, or free, and butterfly.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0,	0	$\frac{0}{19} = 0$	$\frac{0{,}1}{3} = 0{,}03$
Ed.2	The backstroke causes a good sensation after the execution of intense series of crawl, or free style, and butterfly.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0,	1	$\frac{1}{19} = 0{,}05$	
Ed.3	The backstroke causes a good sensation after the execution of intense series of crawl, or free style, and butterfly.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{19} = 0{,}05$	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	Palavras	27		
S-Base	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	Edições			

Ed.1	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{27} = 0$	
Ed.2	Many said that this was a jump created by the Kenyan athletics' school, but it seems to me that it is a variant of the ~~scissor~~ scissors jump.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{27} = 0{,}07$	$\frac{0{,}14}{3} = 0{,}04$
Ed.3	Many people say ~~said~~ that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{27} = 0{,}07$	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	Palavras	17		
S-Base	Many think that the hook is the blow where the fighter throws his hand from the bottom up.	Edições			
Ed.1	Many think that the hook is the ~~blow~~ hit where the fighter hits ~~throws his hand~~ from the bottom ~~up~~ upwards.	Ins.: 0, Ap.: 2, Sub.: 3, MPos.: 0	5	$\frac{5}{17} = 0{,}29$	
Ed.2	Many think that the hook is the blow where the fighter throws his hand from the bottom ~~up~~ upwards.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{17} = 0{,}05$	$\frac{0{,}45}{3} = 0{,}15$
Ed.3	Many think that the hook is the blow where the fighter ~~throws~~ hits with his hand from the bottom up.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{17} = 0{,}11$	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	Palavras	19		
S-Base	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the Copa America.	Edições			
Ed.1	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the Copa America.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	
Ed.2	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the Copa America.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	$\frac{0{,}1}{3} = 0{,}03$
Ed.3	Messi's ~~companion~~ team mate in Barcelona, the midfielder is the Brazilian player with the highest value in the Copa America.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{19} = 0{,}1$	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	Palavras	19		
S-Base	Five series of combined strokes of suple, fireman and ground stabilization with an emphasis on precision of movement.	Edições			
Ed.1	Five series of combined ~~strokes~~ attacks of ~~suple~~ suplex, fireman and ground stabilization with an emphasis on precision of movement.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{19} = 0{,}1$	
Ed.2	Five series of combined strokes of ~~suple~~ suplex, fireman and ~~ground stabilization~~ pin down with an emphasis on precision of movement.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{19} = 0{,}15$	$\frac{0{,}51}{3} = 0{,}17$
Ed.3	Five series of combined ~~strokes~~ moves of ~~suple~~ suplex, fireman carry's slam and ground stabilization with an emphasis on precision of movement.	Ins.: 3, Ap.: 0, Sub.: 2, MPos.: 0	5	$\frac{5}{19} = 0{,}26$	
Ref.	In rugby, the fly-half is the most skilled player in the team.	Palavras	12		
S-Base	In rugby, the aperture is the most skilled player on the team.	Edições			

Ed.1	In rugby, the fly-half ~~aperture~~ is the most ~~skilled~~ talented player on the team.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{12} = 0{,}16$	$\frac{0{,}57}{3} = 0{,}19$
Ed.2	In rugby, the ~~aperture~~ fly-half is the most skilled player in ~~on~~ the team.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{12} = 0{,}16$	
Ed.3	In rugby, the ~~aperture~~ fly-half is the most skilled player ~~on the~~ in a team.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{12} = 0{,}25$	
Ref.	The server is the player who puts the ball into play for the first point.	Palavras	15		
S-Base	The server is the player who puts the ball in play for the first point.	Edições			
Ed.1	The server is the player who puts the ball in play for the first point.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	$\frac{0{,}06}{3} = 0{,}02$
Ed.2	The server is the player who puts the ball in play for the first point.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ed.3	The server is the player who ~~puts~~ hits the ball in play for the first point.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{15} = 0{,}06$	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	Palavras	23		
S-Base	The dream of Tristan Garcia, age 14 and a basketball fan, was to throw a ball into the basket on the school court.	Edições			
Ed.1	The dream of Tristan Garcia, age 14 and a basketball fan, was to throw a ball into the basket on the school court.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	$\frac{0{,}25}{3} = 0{,}08$
Ed.2	The dream of Tristan Garcia, ~~age~~ aged 14 and a basketball fan, was to throw a ball into the basket on the school court.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0{,}04$	
Ed.3	The dream of Tristan Garcia, a basketball fan ~~age~~ aged 14 ~~and a basketball fan~~, was to throw a ball into the basket on the school court.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 3	5	$\frac{5}{23} = 0{,}21$	
Ref.	A bow is for individual and personal use.	Palavras	8		
S-Base	A bow is individual and personal equipment.	Edições			
Ed.1	A bow is individual and personal equipment.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0{,}24}{3} = 0{,}08$
Ed.2	A bow is an individual and personal equipment.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{8} = 0{,}12$	
Ed.3	A bow is an individual and personal equipment.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{8} = 0{,}12$	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	Palavras	21		
S-Base	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	Edições			
Ed.1	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	$\frac{0{,}13}{3} = 0{,}04$
Ed.2	Invest in a good glove, the glove is the ~~biggest~~ greatest safety equipment in boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{21} = 0{,}04$	
Ed.3	Invest in a good glove, the glove is the biggest safety equipment in boxing~~,~~ and it can prevent you from getting seriously hurt.	Ins.: 1, Ap.: 1, Sub.: 0, MPos.: 0	2	$\frac{2}{21} = 0{,}09$	
Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	Palavras	13		
S-Base	All gymnasts competing in the competition jump on a slightly inclined device called a table.	Edições			
Ed.1	All gymnasts competing in the competition jump on a slightly inclined device called a table.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{13} = 0$	$\frac{0{,}53}{3} = 0{,}17$

Ed.2	All gymnasts competing in the competition jump on a slightly inclined device called a ~~table~~ vault.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0,07$	
Ed.3	All gymnasts competing ~~in the competition~~ have to jump on a slightly inclined device called a ~~table~~ vault.	Ins.: 2, Ap.: 3, Sub.: 1, MPos.: 0	6	$\frac{6}{13} = 0,46$	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	Palavras	21		
S-Base	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Edições			
Ed.1	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the ~~apparatus~~ kits.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{21} = 0,04$	
Ed.2	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	$\frac{0,08}{3} = 0,02$
Ed.3	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus family.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{21} = 0,04$	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	Palavras	12		
S-Base	The ribbon is considered the most plastic and characteristic apparatus of gymnastics.	Edições			
Ed.1	The ribbon is considered the most plastic and characteristic apparatus of gymnastics.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{12} = 0$	
Ed.2	The ribbon is considered the most plastic and characteristic apparatus of gymnastics.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{12} = 0$	$\frac{0,08}{3} = 0,02$
Ed.3	The ribbon is considered the most plastic and characteristic apparatus ~~of~~ in gymnastics.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{12} = 0,08$	
Ref.	She won the competition on the apparatus: ribbon and clubs.	Palavras	10		
S-Base	She won the disputes on the devices: ribbon and apples.	Edições			
Ed.1	She won the disputes on the devices: ribbon and clubs ~~apples~~.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{10} = 0,1$	
Ed.2	She won the disputes on the devices: ribbon and clubs ~~apples~~.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{10} = 0,1$	$\frac{0,4}{3} = 0,13$
Ed.3	She won the disputes on the ~~devices~~ apparatus: ribbon and ~~apples~~ clubs.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{10} = 0,2$	
Ref.	The vaulting pole is a very advanced equipment.	Palavras	8		
S-Base	The jumping pole is very advanced equipment.	Edições			
Ed.1	The jumping pole is a very advanced equipment.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{8} = 0,12$	
Ed.2	The ~~jumping~~ vaulting pole is a very advanced equipment.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{8} = 0,25$	$\frac{0,74}{3} = 0,24$
Ed.3	The ~~jumping~~ vaulting pole is a very advanced ~~equipment~~ apparatus.	Ins.: 1, Ap.: 0, Sub.: 2, MPos.: 0	3	$\frac{3}{8} = 0,37$	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	Palavras	23		
S-Base	The striker also made a brilliant move by applying a sheet to an opponent, a move that drew the attention of other players on the web.	Edições			
Ed.1	The striker also made a brilliant move by applying a ~~sheet~~ lob to an opponent, a	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0,04$	$\frac{0,29}{3} = 0,09$

	move that drew the attention of other players on the web.				
Ed.2	The striker also made a brilliant move by ~~applying a sheet to~~ lobbing it over an opponent, a move that drew the attention of other players on the web.	Ins.: 1, Ap.: 2, Sub.: 2, MPos.: 0	5	$\frac{5}{23} = 0{,}21$	
Ed.3	The striker also made a brilliant move by applying a ~~sheet~~ lob to an opponent, a move that drew the attention of other players on the web.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0{,}04$	
Ref.	The board is the most important equipment to catch big waves.	Palavras	11		
S-Base	The board is the most important equipment for catching big waves.	Edições			
Ed.1	The board is the most important equipment for catching big waves.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0,	0	$\frac{0}{11} = 0$	$\frac{0{,}09}{3} = 0{,}03$
Ed.2	The board is the most important equipment for catching big waves.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0,	0	$\frac{0}{11} = 0$	
Ed.3	The board is the most important equipment for ~~catching~~ surfing big waves.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0,	1	$\frac{1}{11} = 0{,}09$	
HTER Total Médio S-Base = $\frac{6{,}9}{50}$ = 0,138 (13.80)					

Fonte: Compilado pelo autor (2020).

APÊNDICE J – AVALIAÇÃO DE TM HTER – S-PRÉ

Tabela 18 – Avaliação de TM HTER - Edições Humanas feitas nas traduções do Sistema de TM Enriquecido Semanticamente e com Injeção Terminológica no Pré-processamento (S-Pré)

Referência / Editores	**Tradução de Referência (Quantidade de Palavras) e Edições das traduções S-Pré**	**Quantidade de palavras / edições**		**HTER Scores**	**HTER Score Médio**
Ref.	A runner does not try to run a marathon in the first days of training.	Palavras	15		
S-Pré	A racer does not try to run a marathon during the first days of training.	Edições			
Ed.1	A racer does not try to run a marathon during the first days of training.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	$\frac{0}{3} = 0$
Ed.2	A racer does not try to run a marathon during the first days of training.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ed.3	A racer does not try to run a marathon during the first days of training.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{1}{15} = 0$	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	Palavras	15		
S-Pré	The thrower is disqualified if he leaves the throwing zone before, during or after the throwing.	Edições			
Ed.1	The thrower is disqualified if he/she leaves the throwing zone before, during or after the throwing.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{15} = 0{,}06$	$\frac{0{,}19}{3} = 0{,}06$
Ed.2	The ~~thrower~~ athlete is disqualified if he leaves the throwing zone before, during or after the ~~throwing~~ throw.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{15} = 0{,}13$	
Ed.3	The thrower is disqualified if he leaves the throwing zone before, during or after the throwing.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	Palavras	18		
S-Pré	Judge Mário Yamasaki decided to stop the combat thinking the fighter had passed out.	Edições			
Ed.1	The referee ~~Judge~~ Mário Yamasaki decided to stop the combat thinking the fighter had passed out.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{18} = 0{,}11$	$\frac{0{,}21}{3} = 0{,}07$
Ed.2	~~Judge~~ Referee Mário Yamasaki decided to stop the combat thinking the fighter had passed out.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{18} = 0{,}05$	
Ed.3	~~Judge~~ Referee Mário Yamasaki decided to stop the combat thinking the fighter had passed out.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{18} = 0{,}05$	
Ref.	The sticking point at which the setter performs the lift.	Palavras	10		
S-Pré	I place the point at which the lifter performs the facelift.	Edições			
Ed.1	I place the point at which the lifter performs the ~~facelift~~ lift.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{10} = 0{,}1$	$\frac{0{,}8}{3} = 0{,}26$
Ed.2	~~I place~~ It sticks the point at which the lifter performs the ~~facelift~~ lift.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{10} = 0{,}3$	
Ed.3	~~I place~~ It was placed the point at which the lifter performs the ~~facelift~~ lift.	Ins.: 1, Ap.: 0, Sub.: 3, MPos.: 0	4	$\frac{4}{10} = 0{,}4$	
Ref.	The winger is the player with less time to think about setting up a strike.	Palavras	15		
S-Pré	The wing is the player that has less time to think in the setup of a play.	Edições			
Ed.1	The wing player is the player that has less time to think in the setup of a play.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{15} = 0{,}06$	$\frac{0{,}32}{3} = 0{,}10$

Ed.2	The ~~wing~~ winger is the player that has less time to think in the setup of a ~~play~~ move.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{15} = 0{,}13$	
Ed.3	The ~~wing~~ winger is the player that has less time to think in the setup of a ~~play~~ move.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{15} = 0{,}13$	
Ref.	The gym has a court on which futsal and handball games can be played.	Palavras	14		
S-Pré	The gym has a court that can host a game of futsal and handball.	Edições			
Ed.1	The gym has a court ~~that can host a game of~~ for futsal and handball.	Ins.: 0, Ap.: 5, Sub.: 1, MPos.: 0	6	$\frac{6}{14} = 0{,}42$	$\frac{0{,}42}{3} = 0{,}14$
Ed.2	The gym has a court that can host a game of futsal and handball.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0,	0	$\frac{0}{14} = 0$	
Ed.3	The gym has a court that can host a game of futsal and handball.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{14} = 0$	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	Palavras	13		
S-Pré	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	Edições			
Ed.1	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{13} = 0$	$\frac{0{,}14}{3} = 0{,}04$
Ed.2	OG Kyle ~~long~~ Long suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0{,}07$	
Ed.3	OG Kyle ~~long~~ Long suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0{,}07$	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	Palavras	23		
S-Pré	In addition, the net has 1.55 cm being higher than that used in tennis and smaller than the net of a volleyball court.	Edições			
Ed.1	In addition, the net ~~has~~ is 1.55 cm, being higher than that used in tennis and smaller than the net of a volleyball court.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{23} = 0{,}08$	$\frac{0{,}47}{3} = 0{,}15$
Ed.2	In addition, the net ~~has~~ being 1.55 cm ~~being~~ higher than that used in tennis and ~~smaller~~ lower than the net of a volleyball court.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 1	3	$\frac{3}{23} = 0{,}13$	
Ed.3	In addition, the net is ~~has~~ 1.55 cm ~~being~~ higher than ~~that~~ the one used in tennis and lower ~~smaller~~ than the net of a volleyball court.	Ins.: 2, Ap.: 2, Sub.: 2, MPos.: 0	6	$\frac{6}{23} = 0{,}26$	
Ref.	The show jumping competition involves a male and female show jumper.	Palavras	11		
S-Pré	The team jump dispute involves a vaulter and a vaulter.	Edições			
Ed.1	The team jump dispute involves ~~a vaulter and a vaulter~~ vaulters of both sexes.	Ins.: 3, Ap.: 4, Sub.: 1, MPos.: 0	8	$\frac{8}{11} = 0{,}72$	$\frac{1{,}8}{3} = 0{,}6$
Ed.2	The team of show ~~jump~~ jumping dispute involves a male ~~vaulter~~ and a female jumper ~~vaulter~~.	Ins.: 4, Ap.: 1, Sub.: 2, MPos.: 0	7	$\frac{7}{11} = 0{,}63$	
Ed.3	The ~~team~~ show jump dispute involves a male ~~vaulter~~ and a female jumper ~~vaulter~~.	Ins.: 2, Ap.: 1, Sub.: 2, MPos.: 0	5	$\frac{5}{11} = 0{,}45$	
Ref.	Tennis, one of the most traditional sports played in the world.	Palavras	11		
S-Pré	Tennis, one of the most traditional and practiced sports in the world.	Edições			
Ed.1	Tennis, one of the most traditional and practiced sports in the world.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	$\frac{0}{3} = 0$
Ed.2	Tennis, one of the most traditional and practiced sports in the world.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	

Ed.3	Tennis, one of the most traditional and practiced sports in the world.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	Palavras	15		
S-Pré	The layup is when the player makes the basket very close to the hoop.	Edições			
Ed.1	The layup is when the player scores ~~makes the basket~~ very close to the hoop.	Ins.: 0, Ap.: 2, Sub.: 1, MPos.: 0	3	$\frac{3}{15} = 0,2$	$\frac{0,72}{3} = 0,24$
Ed.2	The layup is when the player ~~makes~~ hits the basket ~~very close to~~ into the hoop.	Ins.: 0, Ap.: 2, Sub.: 2, MPos.: 0	4	$\frac{4}{15} = 0,26$	
Ed.3	The layup is when the player ~~makes~~ scores the basket ~~very close to~~ into the hoop.	Ins.: 0, Ap.: 2, Sub.: 2, MPos.: 0	4	$\frac{4}{15} = 0,26$	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	Palavras	22		
S-Pré	During a play, a player can give up to two non-consecutive touches, so the team can only give a total of three touches on a ball.	Edições			
Ed.1	During a play, a player can ~~give up~~ perform to two non-consecutive touches, so the team can only give ~~a total of~~ three touches on a ball.	Ins.: 0, Ap.: 4, Sub.: 1, MPos.: 0	5	$\frac{5}{22} = 0,22$	$\frac{0,44}{3} = 0,14$
Ed.2	During ~~a~~ play, a player can ~~give~~ touch up to two non-consecutive touches, so the team can only give a total of three touches on a ball.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{22} = 0,09$	
Ed.3	During a play, a player can ~~give~~ have up to two non-consecutive touches, ~~so~~ and the team can only ~~give~~ have a total of three touches on a ball.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{22} = 0,13$	
Ref.	Mario Suárez just misses the Schwarzer's goal post.	Palavras	8		
S-Pré	Mario Suárez kicks close to the post of Schwarzer's goal.	Edições			
Ed.1	Mario Suárez ~~kicks~~ shoots close to the ~~post of~~ Schwarzer's goal.	Ins.: 0, Ap.: 2, Sub.: 1, MPos.: 0	3	$\frac{3}{8} = 0,37$	$\frac{0,37}{3} = 0,12$
Ed.2	Mario Suárez kicks close to the post of Schwarzer's goal.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	Mario Suárez kicks close to the post of Schwarzer's goal.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ref.	Sailing has been an Olympic sport since 1900.	Palavras	8		
S-Pré	Sailing has been an Olympic sport since 1900.	Edições			
Ed.1	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0}{3} = 0$
Ed.2	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	Palavras	22		
S-Pré	In the last move of the game, center back Gustavo Gómez tackled the forward Corinthian Jô and made the penalty.	Edições			
Ed.1	In the last move of the game, center back Gustavo Gómez tackled the Corinthians forward Jô ~~forward Corinthian~~ and a penalty was made ~~and made the penalty~~.	Ins.: 1, Ap.: 0, Sub.: 2, MPos.: 4	7	$\frac{7}{22} = 0,31$	$\frac{0,4}{3} = 0,13$
Ed.2	In the last move of the game, center back Gustavo Gómez tackled the Corinthians	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 1	2	$\frac{2}{22} = 0,09$	

	forward ~~Corinthian~~ Jô and made the penalty.				
Ed.3	In the last move of the game, center back Gustavo Gómez tackled the forward Corinthian Jô and made the penalty.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{22} = 0$	
Ref.	Evandro was the top scorer, having scored seven goals.	Palavras	9		
S-Pré	The leading scorer is player Evandro, who scored seven goals.	Edições			
Ed.1	The leading scorer is ~~player~~ Evandro, who scored seven goals.	Ins.: 0, Ap.: 1, Sub.: 0, MPos.: 0	1	$\frac{1}{9} = 0{,}11$	$\frac{0{,}33}{3} = 0{,}11$
Ed.2	The leading scorer is player Evandro, who has scored seven goals.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{9} = 0{,}11$	
Ed.3	The ~~lead~~ leading scorer is player Evandro, who scored seven goals.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{9} = 0{,}11$	
Ref.	Breakaway is a technique used in road cycling.	Palavras	8		
S-Pré	Breakaway is a technique used in road cycling.	Edições			
Ed.1	Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0}{3} = 0$
Ed.2	Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	Palavras	21		
S-Pré	The game always starts with a service, play that, obligatorily, alternates between the participants in each game.	Edições			
Ed.1	The game always starts with a service, play that, obligatorily, alternates between the participants in each game.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	$\frac{0{,}42}{3} = 0{,}14$
Ed.2	The game always starts with a serve ~~service~~, ~~play that~~ which must be~~,~~ ~~obligatorily~~ mandatorily~~,~~ alternated ~~alternates~~ between the participants in each game.	Ins.: 2, Ap.: 3, Sub.: 4, MPos.: 0	9	$\frac{9}{21} = 0$	
Ed.3	The game always starts with a service, play that, obligatorily, alternates between the participants in each game.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	Palavras	19		
S-Pré	Playing in position 3, a winger is the player that most closely matches the two extremes of the position of the basketball.	Edições			
Ed.1	Playing in position 3, a winger is the player that most closely matches the two extremes of the position of the basketball.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	$\frac{0}{3} = 0{,}07$
Ed.2	Playing in position 3, a winger is the player that most closely matches the two extremes of the position of the basketball.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	
Ed.3	Playing in position 3, a winger is the player that most closely matches the two extreme ~~extremes of the~~ ~~position~~ positions of the basketball.	Ins.: 0, Ap.: 2, Sub.: 2, MPos.: 0	4	$\frac{4}{19} = 0{,}21$	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	Palavras	18		
S-Pré	The team played well and Juciely, with a slide, play widely used by the sportsman, made 09-06.	Edições			

Ed.1	The team played well and Juciely, with a slide, a play widely used by the sportsperson ~~sportsman~~, made 09-06.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{18} = 0{,}11$	$\frac{0{,}16}{3} = 0{,}05$
Ed.2	The team played well and Juciely, with a slide, a play widely used by the sportsman, made 09-06.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{18} = 0{,}05$	
Ed.3	The team played well and Juciely, with a slide, play widely used by the sportsman, made 09-06.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	Palavras	28		
S-Pré	In the return game of the decision, supporter of the River stoned the bus of the players of the Boca Juniors, who had no condition to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	Edições			
Ed.1	In the return game of the decision, ~~supporter~~ supporters of the River stoned the bus of the players of the Boca Juniors, who had no condition to enter ~~the field of~~ in the Monumental ~~Stadium of~~ de Nuñez Stadium, in Buenos Aires.	Ins.: 0, Ap.: 3, Sub.: 2, MPos.: 1	6	$\frac{6}{28} = 0{,}21$	$\frac{0{,}49}{3} = 0{,}16$
Ed.2	In the return game of the decision, ~~supporter~~ supporters of ~~the~~ River stoned the bus of ~~the players of the~~ Boca Juniors players, who had no condition to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	Ins.: 0, Ap.: 4, Sub.: 1, MPos.: 1	6	$\frac{6}{28} = 0{,}21$	
Ed.3	In the return game of the decision, ~~supporter~~ supporters of the River stoned the bus of the players of ~~the~~ Boca Juniors, who had no condition to enter the field of the Monumental Stadium of Nuñez, in Buenos Aires.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{28} = 0{,}07$	
Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	Palavras	25		
S-Pré	At 18, the captain of the club was the youngest player in the history of the club to score in an Atletiba.	Edições			
Ed.1	At 18, the captain of the club was the youngest player in the history of the club to score in an Atletiba.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{25} = 0$	$\frac{0}{3} = 0$
Ed.2	At 18, the captain of the club was the youngest player in the history of the club to score in an Atletiba.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{25} = 0$	
Ed.3	At 18, the captain of the club was the youngest player in the history of the club to score in an Atletiba.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{25} = 0$	
Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	Palavras	24		
S-Pré	Then the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the event.	Edições			
Ed.1	Then, the swimming pool has to be divided into ten lanes ~~so that~~ and only the eight internal, less turbulent, are used in the event.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{24} = 0{,}08$	$\frac{0{,}24}{3} = 0{,}08$

Ed.2	Then the swimming pool has to be divided into ten lanes so that only the eight internal~~,~~ less turbulent ones~~,~~ are used in the event.	Ins.: 1, Ap.: 2, Sub.: 0, MPos.: 0,	3	$\frac{3}{24} = 0,12$	
Ed.3	Then the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent ones, are used in the event.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{24} = 0,04$	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	Palavras	29		
S-Pré	The theoretically owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	Edições			
Ed.1	The ~~theoretically~~ owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season, theoretically.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 1	2	$\frac{2}{29} = 0,06$	$\frac{0,12}{3} = 0,04$
Ed.2	The ~~theoretically~~ theoretical owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{29} = 0,03$	
Ed.3	The ~~theoretically~~ theoretical owner of the position is again out of action, this time due to injury, while the reserve is the player who has played the most this season.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{29} = 0,03$	
Ref.	The crossing is our strongest skill, like all the teams.	Palavras	10		
S-Pré	The cross is a strong play for us, as it is for all teams.	Edições			
Ed.1	The crossing ~~cross~~ is a strong play for us, as it is for all teams.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{10} = 0,1$	$\frac{0,4}{3} = 0,13$
Ed.2	The crossing ~~cross~~ is a strong ~~play~~ move for us, as it is for all the teams.	Ins.: 1, Ap.: 0, Sub.: 2, MPos.: 0	3	$\frac{3}{10} = 0,3$	
Ed.3	The cross is a strong play for us, as it is for all teams.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	Palavras	18		
S-Pré	However, the sports media usually refer to play only as a scissor kick, with little use of the prefix kick or kicking.	Edições			
Ed.1	However, the sports media usually refer to play only as a scissor kick, without the ~~with little~~ use of the prefix kick or kicking.	Ins.: 1, Ap.: 1, Sub.: 1, MPos.: 0	3	$\frac{3}{18} = 0,16$	$\frac{0,59}{3} = 0,19$
Ed.2	However, the sports media usually refer to ~~play~~ it only as ~~a scissor kick~~ bicycle, with little use of the prefix kick or ~~kicking~~ shot.	Ins.: 0, Ap.: 2, Sub.: 3, MPos.: 0	5	$\frac{5}{18} = 0,27$	
Ed.3	However, the sports media usually refer to play only as a ~~scissor kick~~ bicycle, with little use of the prefix kick or ~~kicking~~ shot.	Ins.: 0, Ap.: 1, Sub.: 2, MPos.: 0	3	$\frac{3}{18} = 0,16$	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	Palavras	17		
S-Pré	The stadium has a nine lane athletics track, two giant screens and a state-of-the-art wi-fi net.	Edições			

Ed.1	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi ~~net~~.	Ins.: 1, Ap.: 1, Sub.: 0, MPos.: 0	2	$\frac{2}{17} = 0{,}11$	$\frac{0{,}16}{3} = 0{,}05$
Ed.2	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi net.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{17} = 0{,}05$	
Ed.3	The stadium has a nine lane athletics track, two giant screens and a state-of-the-art wi-fi net.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{17} = 0$	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	Palavras	20		
S-Pré	Four countries will compete for the challenge of the four stroke of a swimming, where fly is the most complex stroke to learn.	Edições			
Ed.1	Four countries will compete for the challenge of the four stroke of a swimming, where butterfly ~~fly~~ is the most complex stroke to learn.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{20} = 0{,}05$	$\frac{0{,}25}{3} = 0{,}08$
Ed.2	Four countries will compete ~~for~~ in the challenge of the four ~~stroke~~ strokes of ~~a~~ swimming, where butterfly ~~fly~~ is the most complex stroke to learn.	Ins.: 0, Ap.: 1, Sub.: 3, MPos.: 0	4	$\frac{4}{20} = 0{,}2$	
Ed.3	Four countries will compete for the challenge of the four stroke of a swimming, where fly is the most complex stroke to learn.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{20} = 0$	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	Palavras	20		
S-Pré	According to data from the Footstats, the wing was the player with more certain crosses and more pass for goal en a team.	Edições			
Ed.1	According to data from the Footstats, the wing was the player with more ~~certain~~ correct crosses and more pass for goal ~~en~~ in a team.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{20} = 0{,}1$	$\frac{0{,}8}{3} = 0{,}26$
Ed.2	According to data from the Footstats, the ~~wing~~ winger was the player with ~~more certain~~ the most precise crosses and more ~~pass~~ passes to ~~for~~ goal in ~~en~~ the ~~a~~ team.	Ins.: 1, Ap.: 0, Sub.: 7, MPos.: 0	8	$\frac{8}{20} = 0{,}4$	
Ed.3	According to data from the Footstats, the ~~wing~~ winger was the player with more ~~certain~~ precise crosses and more ~~pass~~ passes to ~~for~~ goal in ~~en~~ the ~~a~~ team.	Ins.: 0, Ap.: 0, Sub.: 6, MPos.: 0	6	$\frac{6}{20} = 0{,}3$	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	Palavras	16		
S-Pré	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	Edições			
Ed.1	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the practitioner.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{16} = 0$	$\frac{0{,}12}{3} = 0{,}04$
Ed.2	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the ~~practitioner~~ swimmer.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{16} = 0{,}06$	
Ed.3	Totally different from the other styles, the breaststroke requires a lot of coordination and technique from the ~~practitioner~~ swimmer.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{16} = 0{,}06$	

Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	Palavras	16		
S-Pré	Pedrinho gives a lob, but does not give sequence in a play, to the twelve minutes of the first time.	Edições			
Ed.1	Pedrinho lobs the rival ~~gives a lob~~, but does not give sequence in a play, to the twelve minutes of the first time.	Ins.: 1, Ap.: 1, Sub.: 2, MPos.: 0	4	$\frac{4}{16} = 0{,}25$	$\frac{0{,}81}{3} = 0{,}27$
Ed.2	Pedrinho ~~gives a lob~~ lobbed the player, but did ~~does~~ not give sequence in a play, to the twelve minutes of the first time.	Ins.: 1, Ap.: 1, Sub.: 3, MPos.: 0	5	$\frac{5}{16} = 0{,}31$	
Ed.3	Pedrinho gives a lob, but does not give sequence in ~~a~~ the play, ~~to the~~ at twelve minutes of the first ~~time~~ half.	Ins.: 0, Ap.: 1, Sub.: 3, MPos.: 0	4	$\frac{4}{16} = 0{,}25$	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	Palavras	14		
S-Pré	The traditional celebration with a dive en a court is alive in memory.	Edições			
Ed.1	The traditional celebration with a dive ~~en~~ on a court is alive in our memory.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{14} = 0{,}14$	$\frac{0{,}42}{3} = 0{,}14$
Ed.2	The traditional celebration with a dive on the ~~en a~~ court is alive in the memory.	Ins.: 1, Ap.: 0, Sub.: 2, MPos.: 0	3	$\frac{3}{14} = 0{,}21$	
Ed.3	The traditional celebration with a dive on ~~en~~ a court is alive in memory.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{14} = 0{,}07$	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	Palavras	18		
S-Pré	The position and alignment of a cage in the dispute field is therefore critical to its safe use.	Edições			
Ed.1	The position and alignment of a cage in the ~~dispute~~ field is, therefore, critical to its safe use.	Ins.: 2, Ap.: 1, Sub.: 0, MPos.: 0	3	$\frac{3}{18} = 0{,}16$	$\frac{0{,}21}{3} = 0{,}07$
Ed.2	The position and alignment of ~~a~~ the cage in the dispute field is therefore critical to its safe use.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{18} = 0{,}05$	
Ed.3	The position and alignment of a cage in the dispute field is therefore critical to its safe use.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	Palavras	19		
S-Pré	The back stroke causes a good sensation after performing intense crawl, or free, and butterfly routines.	Edições			
Ed.1	The ~~back stroke~~ backstroke causes a good ~~sensation~~ feeling after performing intense crawl, or free, and butterfly routines.	Ins.: 0, Ap.: 1, Sub.: 2, MPos.: 0	3	$\frac{3}{19} = 0{,}15$	$\frac{0{,}35}{3} = 0{,}11$
Ed.2	The ~~back stroke~~ backstroke causes a good sensation after performing intense crawl, or free style, and butterfly routines.	Ins.: 1, Ap.: 1, Sub.: 1, MPos.: 0	3	$\frac{3}{19} = 0{,}15$	
Ed.3	The back stroke causes a good sensation after performing intense crawl, or free style, and butterfly routines.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{19} = 0{,}05$	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	Palavras	27		
S-Pré	Many said that this was a jump created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissor jump.	Edições			
Ed.1	Many said that this was a jump created by the Kenyan school of athletics, but it	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{27} = 0$	$\frac{0{,}06}{3} = 0{,}02$

	seems to me that it is a variant of the scissor jump.				
Ed.2	Many said that this was a jump created by the Kenyan school of athletics, but it seems to me that it is a variant of the ~~scissor~~ scissors jump.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{27} = 0{,}03$	
Ed.3	Many said that this was a jump created by the Kenyan school of athletics, but it seems to me that it is a variant of the ~~scissor~~ scissors jump.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{27} = 0{,}03$	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	Palavras	17		
S-Pré	Many think that the hook is the stroke where the fighter throws his hand from the bottom up.	Edições			
Ed.1	Many think that the hook is the ~~stroke~~ punch where the fighter punches from the bottom ~~throws his hand up~~ upwards.	Ins.: 0, Ap.: 3, Sub.: 3, MPos.: 0	6	$\frac{6}{17} = 0{,}35$	
Ed.2	Many think that the hook is the stroke where the fighter throws his hand from the bottom ~~up~~ upwards.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{17} = 0{,}05$	$\frac{0{,}40}{3} = 0{,}13$
Ed.3	Many think that the hook is the stroke where the fighter throws his hand from the bottom up.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{17} = 0$	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	Palavras	19		
S-Pré	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the America Cup.	Edições			
Ed.1	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the America Cup.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	
Ed.2	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the America Cup.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	$\frac{0}{3} = 0$
Ed.3	Messi's companion in Barcelona, the midfielder is the Brazilian player with the highest value in the America Cup.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	
Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	Palavras	19		
S-Pré	Five combined stroke routines of suplex, fireman's carry and hold on the floor with an emphasis on movement accuracy.	Edições			
Ed.1	Five combined ~~stroke~~ blows routines of suplex, fireman's carry and pin down ~~hold on the floor~~ with an emphasis on movement accuracy.	Ins.: 0, Ap.: 2, Sub.: 3, MPos.: 0	5	$\frac{5}{19} = 0{,}26$	
Ed.2	Five combined stroke routines of suplex, fireman's carry slam and pin down ~~hold on the floor~~ with an emphasis on movement accuracy.	Ins.: 1, Ap.: 2, Sub.: 2, MPos.: 0	5	$\frac{5}{19} = 0{,}26$	$\frac{0{,}67}{3} = 0{,}22$
Ed.3	Five combined stroke routines of suplex, fireman's carry and ~~hold on the floor~~ pin down with an emphasis on movement accuracy.	Ins.: 0, Ap.: 2, Sub.: 2, MPos.: 0	4	$\frac{4}{19} = 0{,}15$	
Ref.	In rugby, the fly-half is the most skilled player in the team.	Palavras	12		
S-Pré	In rugby, the fly-half is the most skilled player in the club.	Edições			
Ed.1	In rugby, the fly-half is the ~~most skilled~~ best player in the club.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{12} = 0{,}16$	$\frac{0{,}24}{3} = 0{,}08$
Ed.2	In rugby, the fly-half is the most skilled player in the team ~~club~~.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{12} = 0{,}08$	

Ed.3	In rugby, the fly-half is the most skilled player in the club.	Ins. 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{12} = 0$	
Ref.	The server is the player who puts the ball into play for the first point.	Palavras	15		
S-Pré	The server is the player that puts the ball in game for the first point.	Edições			
Ed.1	The server is the player that puts the ball in game for the first point.	Ins. 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	$\frac{0,06}{3} = 0,02$
Ed.2	The server is the player that puts the ball ~~in~~ into game for the first point.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{15} = 0,06$	
Ed.3	The server is the player that puts the ball in game for the first point.	Ins. 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	Palavras	23		
S-Pré	The dream of Tristan Garcia, aged 14 and a basketball fan, was to throw a ball into a basket from a school court.	Edições			
Ed.1	The dream of Tristan Garcia, aged 14 and a basketball fan, was to throw a ball into a basket from a school court.	Ins. 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	$\frac{0,12}{3} = 0,04$
Ed.2	The dream of Tristan Garcia, aged 14 and a basketball fan, was to throw a ball into a basket from the ~~a~~ school court.	Ins. 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0,04$	
Ed.3	The dream of Tristan Garcia, aged 14 and a basketball fan, was to throw a ball into a basket ~~from a~~ of the school court.	Ins. 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{23} = 0,08$	
Ref.	A bow is for individual and personal use.	Palavras	8		
S-Pré	A hoop is an individual and personal apparatus.	Edições			
Ed.1	A hoop is an individual and personal ~~apparatus~~ gear.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0,12$	$\frac{0,36}{3} = 0,12$
Ed.2	A ~~hoop~~ bow is an individual and personal apparatus.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0,12$	
Ed.3	A bow ~~hoop~~ is an individual and personal apparatus.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0,12$	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	Palavras	21		
S-Pré	Invest in a good glove, the glove is the biggest security guard apparatus of boxing, it can prevent you from getting seriously hurt.	Edições			
Ed.1	Invest in a good glove, the glove is the biggest ~~security guard~~ ~~apparatus~~ safety system of boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 1, Sub.: 2, MPos.: 0	3	$\frac{3}{21} = 0,14$	$\frac{0,32}{3} = 0,10$
Ed.2	Invest in a good glove, the glove is the biggest ~~security guard~~ safety apparatus of boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{21} = 0,09$	
Ed.3	Invest in a good glove, the glove is the biggest ~~security guard~~ safety apparatus of boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{21} = 0,09$	
Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	Palavras	13		
S-Pré	All gymnast competing in the event jump on a slightly inclined apparatus called a vault table.	Edições			
Ed.1	All gymnast competing in the event jump on a slightly inclined ~~apparatus~~ object called a vault table.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0,07$	$\frac{0,37}{3} = 0,12$
Ed.2	All ~~gymnast~~ gymnasts competing in the event jump on a slightly inclined apparatus called a ~~vault~~ table.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{13} = 0,15$	

Ed.3	All ~~gymnast~~ gymnasts competing in the event jump on a slightly inclined apparatus called a ~~vault~~ table.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{13} = 0{,}15$	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	Palavras	21		
S-Pré	If the rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Edições			
Ed.1	If the rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the ~~apparatus~~ gears.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{21} = 0{,}04$	$\frac{0{,}04}{3} = 0{,}01$
Ed.2	If the rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Ins. 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	
Ed.3	If the rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Ins. 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	Palavras	12		
S-Pré	The ribbon is considered the most plastic and characteristic apparatus of a gymnastics.	Edições			
Ed.1	The ribbon is considered the most plastic and characteristic ~~apparatus~~ gear of a gymnastics.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{12} = 0{,}08$	$\frac{0{,}16}{3} = 0{,}05$
Ed.2	The ribbon is considered the most plastic and characteristic apparatus of a gymnastics.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{12} = 0$	
Ed.3	The ribbon is considered the most plastic and characteristic apparatus of ~~a~~ gymnastics.	Ins.: 0, Ap.: 1, Sub.: 0, MPos.: 0	1	$\frac{1}{12} = 0{,}08$	
Ref.	She won the competition on the apparatus: ribbon and clubs.	Palavras	10		
S-Pré	She won the contest on the apparatus: ribbon and club.	Edições			
Ed.1	She won the contest on the apparatus: ribbon and club.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	$\frac{0{,}1}{3} = 0{,}03$
Ed.2	She won the contest on the apparatus: ribbon and ~~club~~ clubs.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{10} = 0{,}1$	
Ed.3	She won the contest on the apparatus: ribbon and club.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	
Ref.	The vaulting pole is a very advanced equipment.	Palavras	8		
S-Pré	The pole for jump is a very advanced apparatus.	Edições			
Ed.1	The pole for jump is a very advanced apparatus.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0{,}37}{3} = 0{,}12$
Ed.2	The vaulting pole for jump is a very advanced apparatus.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{8} = 0{,}12$	
Ed.3	The ~~pole~~ vault for ~~jump~~ jumping is a very advanced apparatus.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{8} = 0{,}25$	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	Palavras	23		
S-Pré	The forward also featured a brilliant move to apply a lob to an adversary, a play that drew the attention of other players on the web.	Edições			
Ed.1	The ~~forward~~ player also featured a brilliant move to apply a lob to an adversary, a play that drew the attention of other players on the web.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0{,}04$	$\frac{0{,}29}{3} = 0{,}09$

Ed.2	The forward also featured a brilliant move to ~~apply a lob to~~ lob an adversary, a play that drew the attention of other players on the web.	Ins.: 0, Ap.: 3, Sub.: 0, MPos.: 1	4	$\frac{4}{23} = 0{,}17$	
Ed.3	The forward also featured a brilliant move to ~~apply~~ perform a lob ~~to~~ over an adversary, a play that drew the attention of other players on the web.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{23} = 0{,}08$	
Ref.	The board is the most important equipment to catch big waves.	Palavras	11		
S-Pré	The board is the most important apparatus for catching large waves.	Edições			
Ed.1	The board is the most important ~~apparatus~~ gear for catching large waves.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{11} = 0{,}09$	$\frac{0{,}09}{3} = 0{,}03$
Ed.2	The board is the most important apparatus for catching large waves.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	
Ed.3	The board is the most important apparatus for catching large waves.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	
HTER Total Médio S-Pré $= \frac{5{,}22}{50} =$ **0,1044 (10.44)**					

Fonte: Compilado pelo autor (2020).

APÊNDICE K – AVALIAÇÃO DE TM HTER – S-PÓS

Tabela 19 – Avaliação de TM HTER - Edições Humanas feitas nas Traduções do Sistema de TM Enriquecido Semanticamente e com Injeção Terminológica na Pós-edição (S-Pós)

Referência / Editores	Tradução de Referência (Quantidade de Palavras) e Edições das traduções S-Pós	Quantidade de palavras e edições		HTER Scores	HTER Score Médio
Ref.	A runner does not try to run a marathon in the first days of training.	Palavras	15		
S-Pós	A runner does not try to run a marathon in the first few days of training.	Edições			
Ed.1	A runner does not try to run a marathon in the ~~first~~ few days of training.	Ins.: 0, Ap.: 1, Sub.: 0, MPos.: 0	1	$\frac{1}{15} = 0,06$	$\frac{0,12}{3} = 0,04$
Ed.2	A runner does not try to run a marathon in the first ~~few~~ days of training.	Ins.: 0, Ap.: 1, Sub.: 0, MPos.: 0	1	$\frac{1}{15} = 0,06$	
Ed.3	A runner does not try to run a marathon in the first few days of training.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ref.	The athlete is disqualified if he/she leaves the circle before, during or after the throw.	Palavras	15		
S-Pós	The thrower is disqualified if he leaves the launch zone before, during or after the release.	Edições			
Ed.1	The thrower is disqualified if he/she ~~he~~ leaves the launch zone before, during or after the ~~release~~ move.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{15} = 0,13$	$\frac{0,53}{3} = 0,17$
Ed.2	The ~~thrower~~ athlete is disqualified if he leaves the ~~launch~~ start zone before, during or after the ~~release~~ throw.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{15} = 0,2$	
Ed.3	The athlete ~~thrower~~ is disqualified if he leaves the ~~launch~~ start zone before, during or after the ~~release~~ throw.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{15} = 0,2$	
Ref.	The referee Mario Yamasaki decided to stop the fight because he thought that the fighter had passed out.	Palavras	18		
S-Pós	Ref Mário Yamasaki decided to suspend the fight thinking that the fighter had passed out.	Edições			
Ed.1	Ref Mário Yamasaki decided to suspend the fight thinking that the fighter had passed out.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	$\frac{0,10}{3} = 0,03$
Ed.2	~~Ref~~ Referee Mário Yamasaki decided to suspend the fight thinking that the fighter had passed out.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{18} = 0,05$	
Ed.3	~~Ref~~ Referee Mário Yamasaki decided to suspend the fight thinking that the fighter had passed out.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{18} = 0,05$	
Ref.	The sticking point at which the setter performs the lift.	Palavras	10		
S-Pós	I place the point at which the setter performs the lift.	Edições			
Ed.1	I place the point at which the setter performs the lift.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	$\frac{0,9}{3} = 0,3$
Ed.2	~~I~~ It is a fixing ~~place the~~ point at which the setter performs the lift.	Ins.: 1, Ap.: 0 Sub.: 3, MPos.: 0	4	$\frac{4}{10} = 0,4$	
Ed.3	~~I~~ ~~place~~ It was placed the point ~~at which~~ where the setter performs the lift.	Ins.: 1, Ap.: 1 Sub.: 3, MPos.: 0	5	$\frac{5}{10} = 0,5$	
Ref.	The winger is the player with less time to think about setting up a strike.	Palavras	15		
S-Pós	The winger is the player who has less time to think about setting up a play.	Edições			
Ed.1	The winger is the player who has less time to think about setting up a play.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	$\frac{0}{3} = 0$
Ed.2	The winger is the player who has less time to think about setting up a play.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	

Ed.3	The winger is the player who has less time to think about setting up a play.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ref.	The gym has a court on which futsal and handball games can be played.	Palavras	14		
S-Pós	The gym has a court that can receive futsal and handball games.	Edições			
Ed.1	The gym has a court that can receive futsal and handball games.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{14} = 0$	$\frac{0,14}{3} = 0,04$
Ed.2	The gym has a court that can ~~receive~~ host futsal and handball games.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{14} = 0,07$	
Ed.3	The gym has a court that can ~~receive~~ host futsal and handball games.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{14} = 0,07$	
Ref.	OG Kyle Long injured his hand during the first quarter against the Saints.	Palavras	13		
S-Pós	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	Edições			
Ed.1	OG Kyle long suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{13} = 0$	$\frac{0,14}{3} = 0,04$
Ed.2	OG Kyle Long ~~long~~ suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0,07$	
Ed.3	OG Kyle Long ~~long~~ suffered a hand injury during the first quarter of the game against the Saints.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0,07$	
Ref.	Also, the net is 1.55 cm higher than the one used in tennis and lower than the net used in the volleyball court.	Palavras	23		
S-Pós	In addition, the net is 1.55 cm taller than the one used in tennis and smaller than the net of the volleyball court.	Edições			
Ed.1	In addition, the net is 1.55 cm~~,~~ taller than the one used in tennis and smaller than the net of the volleyball court.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	$\frac{0,08}{3} = 0,02$
Ed.2	In addition, the net is 1.55 cm taller than the one used in tennis and ~~smaller~~ lower than the net of the volleyball court.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0,04$	
Ed.3	In addition, the net is 1.55 cm taller than the one used in tennis and ~~smaller~~ lower than the net of the volleyball court.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{23} = 0,04$	
Ref.	The show jumping competition involves a male and female show jumper.	Palavras	11		
S-Pós	The team jumping competition involves a jumper and a jumper.	Edições			
Ed.1	The team jumping competition involves jumpers ~~a jumper and a jumper~~ of both sexes.	Ins.: 3, Ap.: 4, Sub.: 1, MPos.: 0	8	$\frac{8}{11} = 0,72$	$\frac{1,62}{3} = 0,54$
Ed.2	The team of show jumping competition involves a ~~jumper~~ male and a female jumper.	Ins.: 4, Ap.: 1, Sub.: 0, MPos.: 0	5	$\frac{5}{11} = 0,45$	
Ed.3	The ~~team~~ show jumping competition involves a male show jumper and a female show jumper.	Ins.: 4, Ap.: 0, Sub.: 1, MPos.: 0	5	$\frac{5}{11} = 0,45$	
Ref.	Tennis, one of the most traditional sports played in the world.	Palavras	11		
S-Pós	Tennis, one of the most traditional and practiced sports in the world.	Edições			
Ed.1	Tennis, one of the most traditional and practiced sports in the world.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	$\frac{0}{3} = 0$
Ed.2	Tennis, one of the most traditional and practiced sports in the world.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	
Ed.3	Tennis, one of the most traditional and practiced sports in the world.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{1}{11} = 0$	
Ref.	The layup is when the player lays the ball off the backboard into the hoop.	Palavras	15		

S-Pós	The layup is when the player makes the basket very close to the ring.	Edições			
Ed.1	The layup is when the player scores ~~makes the basket~~ very close to the ring.	Ins.: 0, Ap.: 2, Sub.: 1, MPos.: 0	3	$\frac{3}{15} = 0{,}2$	$\frac{0{,}32}{3} = 0{,}10$
Ed.2	The layup is when the player ~~makes~~ hits the basket very close to the ring.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{15} = 0{,}06$	
Ed.3	The layup is when the player ~~makes~~ hits the basket very close to the ring.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{15} = 0{,}06$	
Ref.	During play, a player can only touch the ball twice not consecutively, and the team can only touch the ball three times.	Palavras	22		
S-Pós	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	Edições			
Ed.1	During a rally, a player can make up to two non-consecutive hits, so the team can only make ~~a total of~~ three hits on the ball.	Ins.: 0, Ap.: 3, Sub.: 0, MPos.: 0	3	$\frac{3}{22} = 0{,}13$	$\frac{0{,}13}{3} = 0{,}04$
Ed.2	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{22} = 0$	
Ed.3	During a rally, a player can make up to two non-consecutive hits, so the team can only make a total of three hits on the ball.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{1}{22} = 0$	
Ref.	Mario Suárez just misses the Schwarzer's goal post.	Palavras	8		
S-Pós	Mario Suárez shoots close to Schwarzer's goal post.	Edições			
Ed.1	Mario Suárez shoots close to Schwarzer's goal post.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0}{3} = 0$
Ed.2	Mario Suárez shoots close to Schwarzer's goal post.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	Mario Suárez shoots close to Schwarzer's goal post.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ref.	Sailing has been an Olympic sport since 1900.	Palavras	8		
S-Pós	Sailing has been an Olympic sport since 1900.	Edições			
Ed.1	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0}{3} = 0$
Ed.2	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	Sailing has been an Olympic sport since 1900.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ref.	At the end of the game, defender Gustavo Gómez slide tackled the Corinthians center forward Jô and a penalty kick was given.	Palavras	22		
S-Pós	In the last game of the game, defender Gustavo Gómez gave a tackle to Corinthians striker Jô and made the penalty.	Edições			
Ed.1	In the last game of the game, defender Gustavo Gómez gave a tackle to Corinthians striker Jô and a penalty kick was given ~~and made the penalty~~.	Ins.: 2, Ap.: 1, Sub.: 2, MPos.: 0	5	$\frac{5}{22} = 0{,}22$	$\frac{0{,}66}{3} = 0{,}22$
Ed.2	In the last ~~game~~ move of the game, defender Gustavo Gómez slide tackled ~~gave a tackle to~~ Corinthians striker Jô and made the penalty.	Ins.: 0, Ap.: 2, Sub.: 3, MPos.: 0	5	$\frac{5}{22} = 0{,}22$	
Ed.3	In the last ~~game~~ moment of the game, defender Gustavo Gómez slide tackled ~~gave a tackle to~~ Corinthians striker Jô and made the penalty.	Ins.: 0, Ap.: 2, Sub.: 3, MPos.: 0	5	$\frac{5}{22} = 0{,}22$	
Ref.	Evandro was the top scorer, having scored seven goals.	Palavras	9		
S-Pós	The top scorer is the player Evandro, who scored seven goals.	Edições			

Ed.1	The top scorer is the player Evandro, who scored seven ~~goals~~ times.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{9} = 0{,}11$	$\frac{0{,}11}{3} = 0{,}03$
Ed.2	The top scorer is the player Evandro, who scored seven goals.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{9} = 0$	
Ed.3	The top scorer is the player Evandro, who scored seven goals.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{9} = 0$	
Ref.	Breakaway is a technique used in road cycling.	Palavras	8		
S-Pós	Breakaway is a technique used in road cycling.	Edições			
Ed.1	Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0}{3} = 0$
Ed.2	Breakaway is a technique used in road cycling	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	Breakaway is a technique used in road cycling.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ref.	The game always starts with a serve, which must be alternated between the participants at the beginning of a new game.	Palavras	21		
S-Pós	The match always begins with a serve, a move that must alternate between the participants in each match.	Edições			
Ed.1	The match always begins with a serve, a move that must alternate between the participants in each match.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	$\frac{0{,}18}{3} = 0{,}06$
Ed.2	The match always begins with a serve, a move that must be ~~alternate~~ alternated between the participants in each match.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{21} = 0{,}09$	
Ed.3	The match always begins with a serve, a move that must be ~~alternate~~ alternated between the participants in each match.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 0	2	$\frac{2}{21} = 0{,}09$	
Ref.	The small forward, number 3, is the player who comes closer to the two extremes of the basketball positions.	Palavras	19		
S-Pós	Playing in position 3, a winger is the player who comes closest to the two ends of the basketball positions.	Edições			
Ed.1	Playing in position 3, a winger is the player who comes closest to the two extremes ~~ends~~ of the basketball positions.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{19} = 0{,}05$	$\frac{0{,}15}{3} = 0{,}05$
Ed.2	Playing in position 3, a winger is the player who comes ~~closest~~ closer to the two ends of the basketball positions.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{19} = 0{,}05$	
Ed.3	Playing in position 3, a winger is the player who comes ~~closest~~ closer to the two ends of the basketball positions.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.:0	1	$\frac{1}{19} = 0{,}05$	
Ref.	The team played well and Juciely, on a slide, a play used by the athlete, made it 09-06.	Palavras	18		
S-Pós	The team played well and Juciely, with a slide, played by the athlete, made 09-06.	Edições			
Ed.1	The team played well and Juciely, with a slide, a play used very often ~~played~~ by the athlete, made 09-06.	Ins.: 4, Ap.: 0, Sub.: 1, MPos.: 0	5	$\frac{5}{18} = 0{,}27$	$\frac{0{,}49}{3} = 0{,}16$
Ed.2	The team played well and Juciely, with a slide, a play ~~played~~ used by the athlete, made it 09-06.	Ins.: 3, Ap.: 0, Sub.: 1, MPos.: 0	4	$\frac{4}{18} = 0{,}22$	
Ed.3	The team played well and Juciely, with a slide, usually played by the athlete, made 09-06.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	
Ref.	On the decisive second leg, River fans stoned the bus of the Boca Juniors players, who were not able to enter Estádio Monumental de Nuñez, in Buenos Aires.	Palavras	28		
S-Pós	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the	Edições			

	Monumental de Nuñez Stadium in Buenos Aires.				
Ed.1	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter ~~the field of~~ in the Monumental de Nuñez Stadium in Buenos Aires.	Ins.: 0, Ap.: 2, Sub.: 1, MPos.: 0	3	$\frac{3}{28} = 0{,}10$	$\frac{0{,}10}{3} = 0{,}03$
Ed.2	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental de Nuñez Stadium in Buenos Aires.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{28} = 0$	
Ed.3	In the return game of the decision, River fans stoned the bus of Boca Juniors players, who were unable to enter the field of the Monumental de Nuñez Stadium in Buenos Aires.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{28} = 0$	
Ref.	At 18, the captain of the team was the youngest player in the history of the club to score in an Atletiba (Atlético x Curitiba).	Palavras	25		
S-Pós	At 18, the team captain was the youngest player in the club's history to score in an Atletiba.	Edições			
Ed.1	At 18, the team captain was the youngest player in the club's history to score in an Atletiba.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{25} = 0$	$\frac{0}{3} = 0$
Ed.2	At 18, the team captain was the youngest player in the club's history to score in an Atletiba.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{25} = 0$	
Ed.3	At 18, the team captain was the youngest player in the club's history to score in an Atletiba.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{25} = 0$	
Ref.	So the swimming pool has to be divided into ten lanes so that only the eight internal, less turbulent, are used in the relays.	Palavras	24		
S-Pós	So the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	Edições			
Ed.1	So, the pool has to be divided into ten lanes so that only the eight indoor, less turbulent ones, are used in the events.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{24} = 0$	$\frac{0{,}16}{3} = 0{,}05$
Ed.2	So the pool has to be divided into ten lanes so that only the eight ~~indoor,~~ internal less turbulent ones, are used in the events.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{24} = 0{,}08$	
Ed.3	So the pool has to be divided into ten lanes so that only the eight ~~indoor,~~ internal less turbulent ones, are used in the events.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{24} = 0{,}08$	
Ref.	Theoretically, the owner of the position is unable to play, this time due to an injury, while the reserve is the player who has played the most this season.	Palavras	29		
S-Pós	The theoretically owner of the position is again out of action, stays time due to injury, while the reserve is the player who has played the most stays season.	Edições			
Ed.1	The ~~theoretically~~ owner of the position is again out of action, ~~stays~~ the time due to injury, while the reserve is the player who has played the most stays season, theoretically.	Ins.: 1, Ap.: 0, Sub.: 1, MPos.: 1,	3	$\frac{3}{29} = 0{,}10$	
Ed.2	The theoretical ~~theoretically~~ owner of the position is again out of action, ~~stays~~ this time due to injury, while the reserve is the player who has played the most ~~stays~~ this season.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{29} = 0{,}10$	$\frac{0{,}30}{3} = 0{,}1$

Ed.3	The theoretical ~~theoretically~~ owner of the position is again out of action, ~~stays~~ this time due to injury, while the reserve is the player who has played the most ~~stays~~ this season.	Ins.: 0, Ap.: 0, Sub.: 3, MPos.: 0	3	$\frac{3}{29} = 0{,}10$	
Ref.	The crossing is our strongest skill, like all the teams.	Palavras	10		
S-Pós	Crossing is a strong move for us, as it is for all teams.	Edições			
Ed.1	Crossing is a strong move for us, as it is for all teams.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	$\frac{0{,}2}{3} = 0{,}06$
Ed.2	Crossing is a strong move for us, as it is for all the teams.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{10} = 0{,}1$	
Ed.3	Crossing is a strong move for us, as it is for all the teams.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{10} = 0{,}1$	
Ref.	However, the sports media usually refers to it as a bicycle, not using the prefix kick or shot.	Palavras	18		
S-Pós	However, the sports media usually refer to the play only as a bicycle kick, with little use of the prefix kick or kick.	Edições			
Ed.1	However, the sports media usually refer to the play only as a bicycle kick, without using ~~with little use of~~ the prefix ~~kick~~ shoot or kick.	Ins.: 0, Ap.: 2, Sub.: 3, MPos.: 0	5	$\frac{5}{18} = 0{,}27$	$\frac{0{,}51}{3} = 0{,}17$
Ed.2	However, the sports media usually refer to the play only as a bicycle ~~kick~~, with little use of the prefix kick or shot ~~kick~~.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{18} = 0{,}12$	
Ed.3	However, the sports media usually refer to the play only as a bicycle ~~kick~~, with little use of the prefix kick or ~~kick~~ shot.	Ins.: 0, Ap.: 1 Sub.: 1, MPos.: 0,	2	$\frac{2}{18} = 0{,}12$	
Ref.	The stadium has an athletics track of nine lanes, two giant screens and a state-of-the-art wi-fi network.	Palavras	17		
S-Pós	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Edições			
Ed.1	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{17} = 0$	$\frac{0}{3} = 0$
Ed.2	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{17} = 0$	
Ed.3	The stadium has a nine-lane athletics track, two giant screens and a state-of-the-art wi-fi network.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{17} = 0$	
Ref.	Four countries will compete in the challenge of four swimming strokes, where butterfly is the most complex stroke to learn.	Palavras	20		
S-Pós	Four countries will compete in the challenge of the four swimming styles, where butterfly is the most complex swim to learn.	Edições			
Ed.1	Four countries will compete in the challenge of the four swimming styles, where butterfly is the most complex swim to learn.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{20} = 0$	$\frac{0{,}1}{3} = 0{,}03$
Ed.2	Four countries will compete in the challenge of the four swimming styles, where butterfly is the most complex ~~swim~~ stroke to learn.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{20} = 0{,}05$	
Ed.3	Four countries will compete in the challenge of the four swimming styles, where butterfly is the most complex stroke ~~swim~~ to learn.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{20} = 0{,}05$	
Ref.	According to data from Footstats, the winger was the player with the best crosses and more assists in the team.	Palavras	20		
S-Pós	According to data from Footstats, the forward was the player with the most correct	Edições			

	crosses and the most passing for goals in the team.				
Ed.1	According to data from Footstats, the forward was the player with the most correct crosses and the most passing for goals in the team.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{20} = 0$	$\frac{0,3}{3} = 0,1$
Ed.2	According to data from Footstats, the forward was the player with the most correct crosses and the most ~~passing for~~ passes to goals in the team.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{20} = 0,1$	
Ed.3	According to data from Footstats, the forward was the player with the most correct crosses and ~~the most passing for~~ more passes to goals in the team.	Ins.: 0, Ap.: 1, Sub.: 3, MPos.: 0	4	$\frac{4}{20} = 0,2$	
Ref.	Totally different from other strokes, breaststroke requires a lot of coordination and technique from the swimmer.	Palavras	16		
S-Pós	Totally different from other strokes, breast stroke requires a lot of coordination and technique from the practitioner.	Edições			
Ed.1	Totally different from other strokes, ~~breast stroke~~ the breaststroke requires a lot of coordination and technique from the practitioner.	Ins.: 1, Ap.: 1, Sub.: 1, MPos.: 0	3	$\frac{3}{16} = 0,18$	$\frac{0,3}{3} = 0,1$
Ed.2	Totally different from other strokes, breast stroke requires a lot of coordination and technique from the ~~practitioner~~ swimmer.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{16} = 0,06$	
Ed.3	Totally different from other strokes, breast stroke requires a lot of coordination and technique from the ~~practitioner~~ swimmer.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{16} = 0,06$	
Ref.	Pedrinho lobs the player, but does not follow through, at twelve minutes of the first half.	Palavras	16		
S-Pós	Pedrinho gives a lob, but does not give sequence in the play, to the twelve minutes of the first half.	Edições			
Ed.1	Pedrinho lobs the player ~~gives a lob~~, but does not give sequence in the play, to the twelve minutes of the first half.	Ins.: 1, Ap.: 1, Sub.: 2, MPos.: 0	4	$\frac{4}{16} = 0,25$	$\frac{0,87}{3} = 0,29$
Ed.2	Pedrinho ~~gives a lob~~ lobbed the player, but did ~~does~~ not give sequence in the play, at ~~to the~~ twelve minutes of the first half.	Ins.: 1, Ap.: 2, Sub.: 4, MPos.: 0	7	$\frac{4}{16} = 0,25$	
Ed.3	Pedrinho ~~gives a lob~~ lobbed, but did ~~does~~ not give sequence in the play, at ~~to the~~ twelve minutes of the first half.	Ins.: 0, Ap.: 3, Sub.: 3, MPos.: 0	6	$\frac{6}{16} = 0,37$	
Ref.	The traditional celebration with a dive on the court is alive in the memory.	Palavras	14		
S-Pós	The traditional celebration with a dive on the court is alive in memory.	Edições			
Ed.1	The traditional celebration with a dive on the court is alive in the memory.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{14} = 0,07$	$\frac{0,21}{3} = 0,07$
Ed.2	The traditional celebration with a dive on the court is alive in the memory.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{14} = 0,07$	
Ed.3	The traditional celebration with a dive on the court is alive in the memory.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{14} = 0,07$	
Ref.	The position and alignment of the cage in the competition field is, therefore, critical for their safe use.	Palavras	18		
S-Pós	The position and alignment of the cage on the competition field is therefore critical to its safe use.	Edições			
Ed.1	The position and alignment of the cage on the ~~competition~~ field is, therefore, critical to its safe use.	Ins.: 2, Ap.: 1, Sub.: 0, MPos.: 0	3	$\frac{3}{18} = 0,16$	$\frac{0,16}{3} = 0,05$
Ed.2	The position and alignment of the cage on the competition field is therefore critical to its safe use.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	

Ed.3	The position and alignment of the cage on the competition field is therefore critical to its safe use.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{18} = 0$	
Ref.	The backstroke causes a good feeling after the execution of intense series of crawl, or free style, and butterfly.	Palavras	19		
S-Pós	The backstroke causes a good sensation after the execution of intense routine of freestyle, or free, and butterfly.	Edições			
Ed.1	The backstroke causes a good sensation after the execution of intense routine of freestyle, or ~~free~~ crawl~~,~~ and butterfly.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{19} = 0{,}1$	$\frac{0{,}25}{3} = 0{,}08$
Ed.2	The backstroke causes a good sensation after the execution of intense routine of freestyle, ~~or free~~ crawl, and butterfly.	Ins.: 0, Ap.: 1, Sub.: 1, MPos.: 0	2	$\frac{2}{19} = 0{,}1$	
Ed.3	The backstroke causes a good sensation after the execution of intense routine of freestyle, or ~~free~~ crawl, and butterfly.	Ins.: 0, Ap.: 0 Sub.: 1, MPos.: 0	1	$\frac{1}{19} = 0{,}05$	
Ref.	Many said that this jump was created by the Kenyan school of athletics, but it seems to me that it is a variant of the scissors jump.	Palavras	27		
S-Pós	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	Edições			
Ed.1	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the scissor jump.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{27} = 0$	$\frac{0{,}06}{3} = 0{,}02$
Ed.2	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the ~~scissor~~ scissors jump.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{27} = 0{,}03$	
Ed.3	Many said that this was a jump created by the Kenyan athletics school, but it seems to me that it is a variant of the ~~scissor~~ scissors jump.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{27} = 0{,}03$	
Ref.	Many people think that the hook is the punch where the fighter punches from the bottom upwards.	Palavras	17		
S-Pós	Many think that the hook is the blow where the fighter throws his hand from the bottom up.	Edições			
Ed.1	Many think that the hook is the ~~blow~~ move where the fighter hits ~~throws his hand~~ from the bottom upwards ~~up~~.	Ins.: 0, Ap.: 2, Sub.: 3, MPos.: 0	5	$\frac{5}{17} = 0{,}29$	$\frac{0{,}39}{3} = 0{,}13$
Ed.2	Many think that the hook is the blow where the fighter throws his hand from the bottom ~~up~~ upwards.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{17} = 0{,}05$	
Ed.3	Many think that the hook is the blow where the fighter throws his hand from the bottom ~~up~~ upwards.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{17} = 0{,}05$	
Ref.	Messi's team mate in Barcelona, the midfielder is the highest valued Brazilian player to play in the Copa América.	Palavras	19		
S-Pós	Messi's companion at Barcelona, the midfielder is the Brazilian player with the highest value to play in the America Cup.	Edições			
Ed.1	Messi's teammate ~~companion~~ ~~at~~ in Barcelona, the midfielder is the Brazilian player with the highest value to play in the America Cup.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{19} = 0{,}10$	$\frac{0{,}15}{3} = 0{,}05$
Ed.2	Messi's companion at Barcelona, the midfielder is the Brazilian player with the highest value to play in the America's Cup.	Ins.: 1, Ap.: 0, Sub.: 0, MPos.: 0	1	$\frac{1}{19} = 0{,}05$	
Ed.3	Messi's companion at Barcelona, the midfielder is the Brazilian player with the highest value to play in the America Cup.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{19} = 0$	

Ref.	Five combined series of blows of suplex, fireman carry's slam and pin down with emphasis on precision of movement.	Palavras	19		
S-Pós	Five routine of combined strokes of suplex, fireman's carry and floor hold with an emphasis on precision of movement.	Edições			
Ed.1	Five routine of combined strokes of suplex, fireman's carry and pin down ~~floor hold~~ with an emphasis on precision of movement.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{19} = 0{,}10$	$\frac{0{,}46}{3} = 0{,}15$
Ed.2	Five routine of combined strokes of suplex, fireman's carry slam and pin down ~~floor hold~~ with an emphasis on precision of movement.	Ins.: 1, Ap.: 0, Sub.: 2, MPos.: 0	3	$\frac{3}{19} = 0{,}15$	
Ed.3	Five ~~routine~~ series of combined strokes of suplex, fireman's carry and ~~floor hold~~ pin down with an emphasis on the precision of movement.	Ins.: 1, Ap.: 0, Sub.: 3, MPos.: 0	4	$\frac{4}{19} = 0{,}21$	
Ref.	In rugby, the fly-half is the most skilled player in the team.	Palavras	12		
S-Pós	In rugby, the opener is the most skilled player on the team.	Edições			
Ed.1	In rugby, the fly-half ~~opener~~ is the most ~~skilled~~ clever player on the team.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{12} = 0{,}16$	$\frac{0{,}4}{3} = 0{,}13$
Ed.2	In rugby, the ~~opener~~ fly-half is the most skilled player in ~~on~~ the team.	Ins.: 0, Ap.: 0, Sub.: 2, MPos.: 0	2	$\frac{2}{12} = 0{,}16$	
Ed.3	In rugby, the fly-half ~~opener~~ is the most skilled player on the team.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{12} = 0{,}08$	
Ref.	The server is the player who puts the ball into play for the first point.	Palavras	15		
S-Pós	The server is the player who puts the ball in game for the first point.	Edições			
Ed.1	The server is the player who puts the ball in game for the first point.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	$\frac{0}{3} = 0$
Ed.2	The server is the player who puts the ball in game for the first point.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ed.3	The server is the player who puts the ball in game for the first point.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{15} = 0$	
Ref.	The dream of Tristan Garcia, aged 14 and a fan of basketball was throwing a ball into the basket of the school court.	Palavras	23		
S-Pós	The dream of Tristan Garcia, a 14-year-old basketball fan, was to throw a ball into the basket on the school court.	Edições			
Ed.1	The dream of Tristan Garcia, a 14-year-old basketball fan, was to throw a ball into the basket on the school court.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	$\frac{0}{3} = 0$
Ed.2	The dream of Tristan Garcia, a 14-year-old basketball fan, was to throw a ball into the basket on the school court.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	
Ed.3	The dream of Tristan Garcia, a 14-year-old basketball fan, was to throw a ball into the basket on the school court.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	
Ref.	A bow is for individual and personal use.	Palavras	8		
S-Pós	A bow is individual and personal equipment.	Edições			
Ed.1	A bow is individual and personal equipment.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0}{3} = 0$
Ed.2	A bow is individual and personal equipment.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ed.3	A bow is individual and personal equipment.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	
Ref.	Invest in a good glove, the glove is the greatest safety equipment in boxing, it can avoid you getting hurt badly.	Palavras	21		
S-Pós	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	Edições			

Ed.1	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	$\frac{0}{3} = 0$
Ed.2	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	
Ed.3	Invest in a good glove, the glove is the biggest safety equipment in boxing, it can prevent you from getting seriously hurt.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	
Ref.	All gymnasts who compete jump on an apparatus slightly inclined called a vault.	Palavras	13		
S-Pós	All gymnasts competing in the competition jump on a slightly inclined equipment called a table.	Edições			
Ed.1	All gymnasts competing in the competition jump on a slightly inclined equipment called a table.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{13} = 0$	$\frac{0,14}{3} = 0,04$
Ed.2	All gymnasts competing in the competition jump on a slightly inclined equipment called a ~~table~~ vault.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0,07$	
Ed.3	All gymnasts competing in the competition jump on a slightly inclined equipment called a vault ~~table~~.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{13} = 0,07$	
Ref.	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment.	Palavras	21		
S-Pós	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Edições			
Ed.1	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the equipment ~~apparatus~~.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{21} = 0,04$	$\frac{0,04}{3} = 0,01$
Ed.2	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	
Ed.3	If rhythmic gymnastics is the black sheep of the gymnastics family, then the rope is the black sheep of the apparatus.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{21} = 0$	
Ref.	The ribbon is considered the most plastic component and characteristic of gymnastics.	Palavras	12		
S-Pós	The ribbon is considered the most plastic and characteristic equipment of gymnastics.	Edições			
Ed.1	The ribbon is considered the most plastic and characteristic equipment of gymnastics.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{12} = 0$	$\frac{0}{3} = 0$
Ed.2	The ribbon is considered the most plastic and characteristic equipment of gymnastics.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{12} = 0$	
Ed.3	The ribbon is considered the most plastic and characteristic equipment of gymnastics.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{12} = 0$	
Ref.	She won the competition on the apparatus: ribbon and clubs.	Palavras	10		
S-Pós	She won the disputes on ribbon and clubs.	Edições			
Ed.1	She won the disputes on ribbon and clubs.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	$\frac{0}{3} = 0$
Ed.2	She won the disputes on ribbon and clubs.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	
Ed.3	She won the disputes on ribbon and clubs.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{10} = 0$	
Ref.	The vaulting pole is a very advanced equipment.	Palavras	8		
S-Pós	The jumping pole is a very advanced equipment.	Edições			
Ed.1	The jumping pole is a very advanced equipment.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{8} = 0$	$\frac{0,24}{3} = 0,08$
Ed.2	The vaulting ~~jumping~~ pole is a very advanced equipment.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0,12$	

Ed.3	The ~~jumping~~ vaulting pole is a very advanced equipment.	Ins.: 0, Ap.: 0, Sub.: 1, MPos.: 0	1	$\frac{1}{8} = 0{,}12$	
Ref.	The center forward also showed brilliance on lobbing the ball over his opponent, which drew the attention of other footballers on the web.	Palavras	23		
S-Pós	The striker also made a brilliant play by applying a lob to an opponent, a play that drew the attention of other players on the web.	Edições			
Ed.1	The striker also made a brilliant play by applying a lob to an opponent, a play that drew the attention of other players on the web.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{23} = 0$	$\frac{0{,}34}{3} = 0{,}11$
Ed.2	The striker also made a brilliant play by ~~applying a lob to~~ lobbing an opponent, a play that drew the attention of other players on the web.	Ins.: 0, Ap.: 3, Sub.: 1, MPos.: 0	4	$\frac{4}{23} = 0{,}17$	
Ed.3	The striker also made a brilliant play by ~~applying a lob to~~ lobbing over an opponent, a play that drew the attention of other players on the web.	Ins.: 0, Ap.: 2, Sub.: 2, MPos.: 0	4	$\frac{4}{23} = 0{,}17$	
Ref.	The board is the most important equipment to catch big waves.	Palavras	11		
S-Pós	The board is the most important equipment for catching big waves.	Edições			
Ed.1	The board is the most important equipment for catching big waves.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	$\frac{0}{3} = 0$
Ed.2	The board is the most important equipment for catching big waves.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	
Ed.3	The board is the most important equipment for catching big waves.	Ins.: 0, Ap.: 0, Sub.: 0, MPos.: 0	0	$\frac{0}{11} = 0$	
HTER Total Médio S-Pós $= \frac{3{,}69}{50}$ **= 0,0738 (7.38)**					

Fonte: Compilado pelo autor (2020).