

UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS



**Avaliação de qualidade de tradução de diálogos
na aplicação didática – *ELSA Speak***

Mariana Rodrigues Almeida

Relatório de estágio orientado pela Professora Doutora Helena Moniz e coorientado pela Doutora Ângela Costa, especialmente elaborado para a obtenção do grau de Mestre em Tradução

2021

Agradecimentos

Este trabalho é dedicado à minha família.

À minha mãe que sempre esteve presente.

À minha irmã que sempre me incentivou.

Aos meus avós que sempre acreditaram em mim.

Aos meus amigos que não me deixaram desistir.

À minha orientadora e coorientadora que sempre me acompanharam nesta aventura, ajudando e apoiando quando mais precisava.

Muito obrigada a todos os que permitiram que este trabalho fosse possível.

Lista de abreviaturas

CBMT – Corpus-based Machine Translation

DQF – Dynamic Quality Framework

EBMT – Example-based Machine Translation

IPA – International Phonetic Alphabet

MQM – Multidimensional Quality Metric

MT – Machine Translation

NMT – Neural Machine Translation

RBMT – Rule-based Machine Translation

SMT – Statistical Machine Translation

TAUS – Translation Automation User Society

TQA – Translation Quality Assessment

Índice de Figuras

Figura 1 - Resultado do Teste de Avaliação	15
Figura 2 - Resultado do Teste de Avaliação com o nível de proficiência do utilizador	15
Figura 3 - Sequência de Planetas da aplicação <i>ELSA Speak</i>	19
Figura 4 - Menu de ajuda do planeta <i>Schwa</i> da aplicação <i>ELSA Speak</i>	19
Figura 5 - Exercício de <i>Pronunciation</i> da aplicação <i>ELSA Speak</i>	20
Figura 6 - Exercício de <i>Listening</i> da aplicação <i>ELSA Speak</i>	20
Figura 7 - Exercício de <i>Word Stress</i> da aplicação <i>ELSA Speak</i>	20
Figura 8 - Exercício de <i>Conversation</i> da aplicação <i>ELSA Speak</i>	20
Figura 9 - Comentário fonético breve	21
Figura 10 - Comentário fonético detalhado	21
Figura 11 - Exercício do dicionário da aplicação <i>ELSA Speak</i>	22
Figura 12 - Comentário detalhado de um exercício do dicionário	22
Figura 13 - Separador Progresso na aplicação <i>ELSA Speak</i>	22
Figura 14 – Tipos de erro do <i>LISA QA Model</i>	33
Figura 15 – Estrutura Hierárquica da MQM/DQF	35
Figura 16 – MQM <i>Core</i>	36
Figura 17 – Categoria <i>Accuracy</i> da MQM	47
Figura 18 - Categoria <i>Fluency</i> da MQM	48

Índice de Tabelas

Tabela 1 – Menus de ajuda da aplicação <i>ELSA Speak</i>	16
Tabela 2 – Fases prototípicas da conversação segundo Moreno (2002)	40
Tabela 3 - Exemplo da base de dados	46
Tabela 4 – Exemplo dos campos extra de análise	46

Índice de Gráficos

Gráfico 1 – Número de erros nos motores de tradução automática	51
Gráfico 2 – Classificação de erros	52
Gráfico 3 – Número de erros nas categorias <i>Accuracy</i> e <i>Fluency</i>	52
Gráfico 4 – Erros na categoria <i>Accuracy</i>	53
Gráfico 5 – Erros na categoria <i>Fluency</i>	62

RESUMO

Este trabalho tem como objetivo averiguar que tipo de erros os motores de tradução Google Tradutor e *DeepL* mais cometem, em contexto de diálogos informais. Para esse efeito, foram analisadas as traduções de 186 exercícios de diálogos da aplicação *ELSA Speak* – uma aplicação que auxilia os seus utilizadores na pronúncia de inglês. Todos os erros encontrados foram catalogados nas tipologias de erros de *Accuracy* e *Fluency* da tipologia de erros da MQM (*Multidimensional Quality Metric*), sendo também classificados com diferentes graus de gravidade: menor, grave e crítico.

Na categoria *Accuracy*, ou adequação do texto de chegada na tradução proposta no texto de partida, foi notório que a maioria dos erros apontados foram provocados por traduções demasiado literais dos textos de partida, tanto do motor de tradução automática Google Tradutor, como do *DeepL*. No entanto, foram também registados alguns erros de omissões e adições de informação, texto que não foi devidamente traduzido ou que não deveria ter sido traduzido de todo. Já na categoria *Fluency*, ou fluência do texto de chegada, os erros mais evidentes estão relacionados com erros de coerência textual, existindo também alguns erros de concordância gramatical, de tempos e modos verbais incorretos, de diacríticos e alguns erros de pontuação. Após a análise dos erros encontrados, concluiu-se que o motor de tradução automática *DeepL* apresenta melhores resultados do que o Google Tradutor.

Será também brevemente apresentada a estrutura global da conversação, assim como todas as fases que a compõem. A nível de microestrutura da conversação, será analisada a tomada de palavra e a sua importância num diálogo.

Palavras-chave: Tradução Automática, Tradução Neuronal, Tipologia de Erros, *Multidimensional Quality Metric*

ABSTRACT

This study investigates what sort of errors the machine translation systems Google Translator and DeepL most frequently produce in a context of informal dialogues. To do so, 186 dialogues from the app ELSA Speak – an app that helps users to improve their English pronunciation – were analyzed. All the errors that were found were categorized according to the Multidimensional Quality Metric error typology, as well as classified with different degrees of severity: minor, major and critical.

In the category Accuracy, it was clear that most errors were motivated by literal translations of the source texts, in both Google Translator and DeepL automatic translated dialogues. However, there were also other types of errors, such as omission or addition of information, and text that should not have been translated or was partly translated. As for the category Fluency, most errors were related to textual coherence, but there were also some grammatical agreement errors, incorrect use of the verb tenses and mood, fewer errors on diacritics and punctuation. After a careful analysis of all these errors, the results pointed out that DeepL currently produces better and cleaner translations when compared with Google Translator.

The overall structure of the dialogue will also be briefly presented, as well as all its phases. At a microstructure level, it will be analyzed the turn-taking and its importance in a dialogue.

Keywords: Machine Translation, Neural Machine Translation, Errors Typologies, Multidimensional Quality Metric

Índice

Agradecimentos.....	1
Lista de abreviaturas	2
Índice de Figuras	3
Índice de Tabelas.....	3
Índice de Gráficos	4
RESUMO	5
ABSTRACT.....	7
Índice.....	9
1. Introdução	11
2. Apresentação da ELSA e tarefas realizadas	13
2.1. Apresentação da ELSA Corp.	13
2.2. Apresentação da aplicação ELSA Speak.....	14
2.3. Tarefas realizadas ao longo do estágio.....	22
3. Estado de arte em Tradução Automática.....	25
3.1. A tradução Automática e o seu desenvolvimento histórico	25
3.2. A Avaliação de Qualidade de Tradução e primeiros esforços de padronização	30
3.3. A tipologia de erros MQM.....	33
3.4. Diálogos Informais	37
4. Metodologia e Corpus	43
4.1. Corpus	43
4.2. Construção da base de dados para a análise de erros de tradução.....	44
4.3. Critérios de anotação de erros	46
5. Resultados e Discussão	50
5.1. Erros mais frequentes	50
5.1.1. Distribuição de erros na categoria <i>Accuracy</i>	52

5.1.1.1.	<i>Overly Literal</i> (Tradução Literal).....	52
5.1.1.2.	<i>Mistranslation of Technical Relationship</i> (Erro de tradução de natureza técnica)..	54
5.1.1.3.	<i>Omission</i> (Omissão):.....	55
5.1.1.4.	<i>Untranslated</i> (Não traduzido):	56
5.1.1.5.	<i>Addition</i> (Adição).....	58
5.1.1.6.	<i>Should Not Have been Translated</i> (Não deveria ter sido traduzido).....	59
5.1.1.7.	<i>Under-Translation</i> (Traduzido de forma incompleta):.....	60
5.1.2.	Distribuição de erros na categoria <i>Fluency</i>	61
5.1.2.1.	<i>Coherence</i> (Coerência).....	62
5.1.2.2.	<i>Agreement</i> (Concordância Gramatical)	63
5.1.2.3.	<i>Tense/Mood/Aspect</i> (Tempos e Modos Verbais).....	65
5.1.2.4.	<i>Word Order</i> (Ordem das palavras).....	66
5.1.2.5.	<i>Diacritics</i> (Diacríticos).....	67
5.1.2.6.	<i>Punctuation</i> (Pontuação)	68
6.	Conclusão e Trabalho futuro	69
	Bibliografia	70

1. Introdução

O presente relatório de estágio é realizado no âmbito do Mestrado de Tradução pela Faculdade de Letras da Universidade de Lisboa e tem como base uma análise dos erros de tradução cometidos pelos motores de Tradução Automática Google Tradutor e *DeepL* em contexto de diálogo informal.

Para a obtenção do grau de mestre, realizei um estágio curricular, de cinco meses, na empresa *Elsa Corp.*, local em que pós-editei as traduções feitas pelo Google Tradutor e pelo *DeepL*, de inglês para português, dos vários exercícios da aplicação *ELSA Speak*. Ao longo do estágio, constatei que ambos os motores de tradução automática cometiam inúmeros erros de tradução, que poderiam ser prejudiciais para a aplicação, uma vez que um dos seus objetivos é que os utilizadores compreendam, na sua língua nativa, todos os exercícios apresentados. A grande maioria dos utilizadores da aplicação *ELSA Speak* pretende treinar a sua pronúncia, para conseguir manter um bom diálogo no seu dia-a-dia, praticando nos diversos exercícios de pronúncia, audição, sílaba tónica e diálogo, disponíveis na aplicação. Por esse motivo, foi considerado importante compreender que tipo de erro estes motores de tradução produzem em contexto dos diálogos informais apresentados na aplicação, lembrando também que estes sistemas de tradução são utilizados diariamente por milhares de utilizadores *online*.

Este relatório de estágio divide-se em cinco capítulos que irei apresentar, brevemente, de seguida. O primeiro capítulo, intitulado “Apresentação da *ELSA* e tarefas realizadas”, divide-se em três secções. Na secção 2.1, será apresentada a empresa na qual realizei o estágio, os seus fundadores e a sua missão. De seguida, na secção 2.2 será divulgada a aplicação *ELSA Speak*, nomeadamente qual o seu principal objetivo e as suas funcionalidades e exercícios. Na última secção deste capítulo, 2.3, serão explicadas as minhas tarefas na empresa durante os cinco meses de estágio.

No segundo capítulo, intitulado “Estado de arte em Tradução Automática”, será apresentada, na secção 3.1, uma breve resenha histórica da Tradução Automática. Na secção 3.2, será abordada a avaliação da qualidade das traduções e os esforços que foram realizados na literatura, para atingir um consenso de como este processo se poderia uniformizar. O que nos leva à secção 3.3, em que será explicada a tipologia de erros MQM, utilizada em todo o mundo para a avaliação de traduções. Por fim, na secção 3.4, serão comentadas algumas

características dos diálogos informais, como as fases prototípicas da conversação e a tomada de turnos.

O terceiro capítulo, “Metodologia e Corpus”, está dividido em 3 secções. Na primeira secção, 4.1, apresentar-se-á uma breve explicação do corpus que me foi fornecido pela empresa, para o cumprimento da minha tarefa principal de pós-edição das traduções do Google Tradutor e do *DeepL*. Na secção 4.2, será primeiramente apresentada a base de dados por mim criada com todos os erros de tradução assinalados e classificados como menores, graves ou críticos, segundo a métrica MQM. De seguida, serão descritos alguns campos extra de análise, também criados por mim, que contêm o número de erros em cada diálogo, assim como o grau de severidade desses mesmos erros e o número de palavras por diálogo. Já na secção 4.3, serão explicados os critérios de anotação dos erros apresentados neste estudo.

No quarto capítulo, nomeado “Resultados e Discussão”, serão apresentados os resultados obtidos a partir da análise dos erros de tradução. Na secção 5.1, serão expostos os erros mais frequentes do Google Tradutor e do *DeepL*, estando na subsecção 5.1.1 os vários erros catalogados na categoria *Accuracy* e na subsecção 5.1.2 os erros da categoria *Fluency*, da métrica MQM.

Por fim, serão apresentadas as minhas conclusões no quinto e último capítulo deste relatório de estágio, denominado “Conclusão e trabalhos futuros”, sendo que também identificarei as minhas contribuições com este estudo.

2. Apresentação da ELSA e tarefas realizadas

Neste capítulo irá ser apresentada, na secção 2.1, a *startup* ELSA Corp, como foi fundada e qual a sua missão. De seguida, na secção 2.2, será feita a descrição da aplicação ELSA *Speak*, que tem como principal objetivo o treino da pronúncia em Inglês Americano, e serão explicadas as suas funcionalidades. Por fim, na secção 2.3, será descrito o meu papel como estagiária, quais as tarefas por mim realizadas e a importância das mesmas no contexto geral da empresa.

2.1. Apresentação da ELSA Corp.

A ELSA Corp. é uma *startup* fundada em 2015 por Vu Van e Xavier Anguera. Em 2015, Vu Van foi trabalhar para os Estados Unidos e, enquanto lá vivia, sentiu que deveria melhorar a sua pronúncia ao ponto de conseguir falar como um falante nativo. Para o conseguir, começou a ter aulas com professores especializados. Foi neste momento que Vu teve a ideia de criar uma aplicação que poderia ajudar outros que, tal como ela, queriam ter uma pronúncia perfeita, mas que não tinham capacidade financeira para recorrer a aulas privadas. Assim, em 2015, a ELSA foi criada. Vu Van é a CEO (*Chief Executive Officer*) da *startup* e Xavier Anguera é o CTO (*Chief Technology Officer*).

A ELSA Corp. está sediada em São Francisco e tem também escritórios em Lisboa, na Índia e no Vietname, país no qual tem a maioria dos seus utilizadores, e já conta com uma equipa de mais de 50 pessoas. Em Lisboa, onde realizei o meu estágio, a equipa de fala é liderada por Xavier Anguera. Esta equipa é composta por cientistas da área de processamento de fala e linguistas. Os engenheiros dedicam-se ao aperfeiçoamento dos modelos de reconhecimento de fala, à melhoria do sistema de avaliação dos utilizadores, à recomendação de conteúdo indicado para cada um dos utilizadores (baseado na dificuldade das lições e no nível de proficiência do utilizador ao produzir os diferentes sons presentes nas lições), entre muitas outras tarefas. Por sua vez, os linguistas avaliam a qualidade do conteúdo lançado semanalmente: avaliando a funcionalidade de conteúdo e a qualidade dos áudios que acompanham os exercícios, verificando as transcrições fonéticas das palavras presentes nos exercícios e avaliando a qualidade das traduções disponibilizadas na aplicação. Os linguistas, para além destas tarefas, trabalham também na criação de recursos de treino e avaliação, nomeadamente na etiquetagem de dados

áudio e texto que permitem avaliar a tecnologia utilizada e melhorá-la. Todos os elementos desta equipa, que estão em constante contacto com pessoas de outros escritórios, são cruciais para o excelente funcionamento da aplicação e certificam-se de que há um controlo efetivo de erros, para que o utilizador possa usufruir da melhor experiência possível.

2.2. Apresentação da aplicação ELSA Speak¹

A ELSA (*English Language Speech Assistant*) é a professora virtual de pronúncia que, usando tecnologia de reconhecimento automático de fala, avalia de forma precisa a pronúncia de inglês do utilizador. Foi adotado o inglês americano visto ser a variante com mais influência e com mais falantes nativos. A aplicação, disponível para android e iOS, já tem mais de 4 milhões de utilizadores em mais de 100 países e permite praticar inglês num ambiente divertido e relaxado. Dispõe de mais de 6000 lições práticas criadas por diversos especialistas de renome mundial na área da pronúncia e comunicação de inglês. A ELSA já ganhou diversos prémios entre os quais o prémio de ouro de melhor aplicação pela *Reimagine Education* de 2016 e foi também, nesse mesmo ano, classificada como aplicação número um no *ProductHunt*.

Nesta aplicação didática, a fonética tem um papel de extrema importância. Como já foi referido anteriormente, o objetivo da ELSA *Speak* é auxiliar os seus utilizadores a serem competentes em termos de pronúncia. Para isso, o utilizador dispõe de baterias de exercícios para conhecer o inventário fonético do inglês americano, os sons mais problemáticos de produzir por um falante nativo e simulação de diálogos, para contextualização dos sons. A ELSA concilia a fonética com o conhecimento prévio do utilizador, o que vai enriquecer o processo de aprendizagem. O alfabeto fonético utilizado na aplicação é o Alfabeto Fonético Internacional – AFI *International Phonetic Alphabet* –IPA).

Quando o utilizador entra na aplicação pela primeira vez, é-lhe recomendado fazer um Teste de Avaliação (*Assessment Test*). O propósito do teste é averiguar qual o seu nível de proficiência de pronúncia de inglês. O utilizador tem apenas de repetir várias frases.

¹ Todos os dados relativos à aplicação ELSA Speak, como a descrição da aplicação e dos seus exercícios, foram recolhidos até à data de setembro de 2019.

Neste teste, a ELSA vai focar-se na pronúncia dos fonemas dessas frases e, de seguida, vai comparar a pronúncia do utilizador com a de uma falante nativa que gravou todo o conteúdo da aplicação, personificando assim a ELSA. O resultado do Teste de Avaliação vai incluir a percentagem de falante nativo (*native speaker*) do utilizador e o seu nível de proficiência, tendo em consideração os fonemas testados em cada frase do teste.

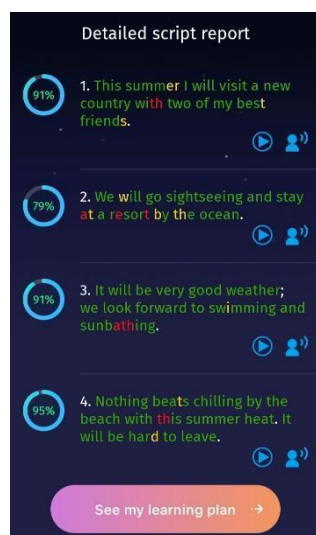


Figura 1 - Resultado do Teste de Avaliação



Figura 2 - Resultado do Teste de Avaliação com o nível de proficiência do utilizador

Depois do Teste de Avaliação realizado, a ELSA vai criar um plano de aprendizagem customizado relativamente ao nível de proficiência do utilizador, de acordo com os resultados obtidos. Este plano de aprendizagem está dividido em *skills* (traduzirei *skills* por competências, como será doravante referido) e cada competência vai testar determinadas classes de sons e tipos de sílabas de maior dificuldade de produção por um não nativo. Estas competências foram agrupadas, na aplicação, em planetas e o critério de seleção para a criação de um planeta é o de determinados sons poderem ser confundidos entre eles pelos utilizadores ou de dificuldades de produção. Os planetas estão organizados de acordo com as dificuldades do utilizador e, se este tiver mais dificuldades numa determinada competência, esse planeta vai aparecer em primeiro e as competências que o utilizador já dominou estarão em último, apenas como reforço. Assim, cada utilizador tem uma sequência de planetas diferente. Esta estratégia de organização permite uma aprendizagem mais rápida pois destaca as competências que o utilizador deve priorizar e praticar em vez de outras que já dominou.

Na aplicação, o utilizador vai encontrar em todos os planetas (no ponto de interrogação, ou seja, na ajuda) toda a informação relativa a esse planeta, os fones pertencentes a cada competência, a razão por estes estarem assim agrupados e ainda alguns erros que os utilizadores podem cometer ao pronunciarem estes fones. Não sendo o foco principal deste relatório, serão apresentados os menus de ajuda dos planetas com o intuito de incluir informação complementar sobre a aplicação *ELSA Speak*. Estes menus de ajuda foram elaborados na perspetiva do utilizador e a sua informação resulta da análise de dados de produção que salientam as categorias com mais problemas. Coligiu-se toda a informação dos menus de ajuda da aplicação na tabela que a seguir se apresenta:

/p/ /t/ /k/	/p/ (<i>pen</i>), /t/ (<i>two</i>) e /k/ (<i>can</i>) são oclusivas que frequentemente no inglês são aspiradas. Isto significa que são tipicamente pronunciadas com uma exalação de ar. Visto que muitas línguas não têm oclusivas aspiradas, um erro comum é pronunciar /p/, /t/ e /k/ sem serem aspirados.
/w/ /v/ /b/	/w/ (<i>west</i>), /v/ (<i>vest</i>) e /b/ (<i>best</i>) são fones labiais e precisam do movimento dos lábios para a sua pronúncia. Há alguns erros possíveis que podem ser cometidos. Alguns falantes dizem /v/ em vez de /w/, assim, <i>vest</i> soa a <i>west</i> . O oposto pode também acontecer, com <i>west</i> a soar a <i>vest</i> . Outros substituem /v/ por /b/, fazendo com que <i>vest</i> soe a <i>best</i> .
/s/ /ʃ/ /z/	/s/ /ʃ/ /z/ são consoantes fricativas sibilantes. /s/ é o primeiro som em <i>sun</i> , /ʃ/ (<i>sh</i>) é o primeiro som em <i>shoe</i> e /z/ é o primeiro som em <i>zero</i> . Algumas línguas não têm estes três sons e, por isso, é comum haver erros. Os sons /s/ e /ʃ/ podem ser confundidos. Consequentemente, cometem o erro de pronunciarem <i>see</i> e <i>she</i> da mesma maneira. Outro erro comum é fazer com que o som /s/ seja pronunciado como /z/, não havendo assim diferença entre <i>loose</i> e <i>lose</i> .
Consonant Clusters (Encontro Consonantal)	Quando duas ou mais consoantes são pronunciadas em sequência, é chamado de encontro consonantal. Estas consoantes podem ser difíceis de pronunciar, especialmente se na língua materna do falante não existem. As palavras <i>blue</i> e <i>Spring</i> começam com encontro consonantal e <i>asked</i> e <i>tests</i> acabam em encontro consonantal. É comum os falantes cometerem o erro de deixarem cair consoantes (dizendo <i>ask</i> em vez de <i>asked</i>) ou adicionarem uma vogal entre as consoantes (dizendo <i>bulue</i> em vez de <i>blue</i>). Muitas palavras do inglês têm encontros consonantais, incluindo a maioria dos verbos no passado (como <i>asked</i>) e a maior parte dos substantivos no plural (como <i>tests</i>) e, por isso, é muito importante praticar.
R sounds: /r/ /ɜ:/ /ə/	Os róticos são sons de difícil aprendizagem. Este é um som único que pode agir como consoante ou vogal. Quando é utilizado como consoante (em <i>red</i>) escreve-se /r/. Quando é utilizada como vogal (em <i>her</i>) escreve-se /ɜ:/ ou /ə/ (por vezes, também se escreve <i>er</i> para este som). O erro mais comum é a substituição de /r/ por /w/ ou outra consoante (dizendo <i>wed</i> em vez de <i>red</i>) ou a alteração da vogal r para uma vogal normal (dizendo <i>ha</i> em vez de <i>her</i>).
/i/ /I/	A vogal na palavra <i>eat</i> escreve-se /i/ (por vezes as pessoas escrevem-na <i>ee</i>) e a vogal na palavra <i>it</i> escreve-se /I/ (que é por vezes escrita como <i>i</i> ou <i>ih</i>). Estes dois sons são

	<p>muito parecidos, mas não são iguais. Muitas línguas têm um dos sons, mas não os dois.</p> <p>É muito comum que os estudantes de inglês pronunciem estes sons da mesma maneira, o que poderá levar a que palavras como <i>live</i> e <i>leave</i> sejam difíceis de diferenciar.</p>
/l/ /r/	<p>(l) (<i>light</i>) e /r/ (<i>right</i>) são consoantes laterais difíceis de pronunciar por muitos utilizadores.</p> <p>Um erro comum é fazer com que /l/ e /r/ soem iguais, o que pode causar alguma confusão aos ouvintes. Outros falantes não têm problemas com o /l/ no início de uma palavra, mas cometem o erro de mudar o /l/ para uma vogal no fim de uma palavra (dizendo algo como <i>ba</i> em vez de <i>ball</i>).</p>
Ending Sounds (Codas consonantais)	<p>As palavras <i>like</i>, <i>had</i> e <i>bag</i> terminam com consoantes.</p> <p>Erros comuns incluem apagamento da consoante (dizendo <i>lie</i> em vez de <i>like</i>), adicionar uma vogal depois da consoante (dizendo <i>hado</i> em vez de <i>had</i>) e mudar o som da consoante (dizendo <i>back</i> em vez de <i>bag</i>).</p>
/ʃ/ /ʒ/ /tʃ/ /dʒ/	<p>/ʃ/ (sh) é o primeiro som em <i>she</i>, /ʒ/ (zh) é a primeira consoante em <i>usual</i>, /tʃ/ (ch) é o primeiro som em <i>chair</i> e /dʒ/ (j) é o primeiro som em <i>just</i>.</p> <p>Um erro comum é fazer com que duas destas consoantes soem iguais ou muito similares. Por exemplo, alguns falantes cometem o erro de dizer <i>wash</i> e <i>watch</i> da mesma maneira, apesar de os sons finais destas palavras não serem iguais.</p>
TH Sounds: /θ/ /ð/	<p>Os sons /θ/ /ð/ merecem um planeta exclusivo, uma vez que a distinção sonoro/surdo é de difícil aquisição na produção de “th”. Existem duas versões do som th: uma delas é surda. É escrito como /θ/ e é o primeiro som em <i>thanks</i>. A outra versão é sonora. Este é escrito como /ð/ e é o primeiro som em <i>the</i>.</p> <p>Existem muitos erros possíveis com os sons th. Os falantes poderão substituí-los com /s/ e /z/ (dizendo <i>sanks</i> em vez de <i>thanks</i>) ou cometer o erro de dizer /f/ e /v/ (dizendo <i>wif</i> em vez de <i>with</i>). Os falantes poderão ainda utilizar /t/ e /d/ (dizendo <i>dare</i> em vez de <i>there</i>).</p>
Schwa: /ə/	<p>A vogal mais comum em inglês é a <i>schwa</i> e o seu símbolo é /ə/ (poderá ser escrito como uh). Este som é apenas encontrado em sílabas átonas. É a primeira vogal nas palavras <i>around</i> e <i>together</i>. A palavra <i>America</i> tem dois <i>schwas</i>: um no início e outro no fim.</p> <p>O som schwa pode ser pronunciado como /e/ na primeira vogal de <i>around</i>, um som forte soando assim a <i>ay-round</i>. Do mesmo modo, alguns falantes cometem o erro de dizer /u/ em vez de /ə/ na primeira vogal de <i>together</i> soando assim a <i>too-gether</i>.</p>
/j/ (y) /ʒ/ /dʒ/	<p>O primeiro som na palavra <i>yes</i> escreve-se /j/ (este é o símbolo do IPA para o som y). A primeira consoante em <i>usual</i> é /ʒ/ (também é escrito como zh). O primeiro som em <i>just</i> escreve-se /dʒ/ (o símbolo do som j). Visto que estes sons são semelhantes, os falantes podem confundir-los.</p> <p>Um erro comum é pronunciar estas três consoantes da mesma maneira ou trocá-las. Por exemplo, alguns falantes utilizam acidentalmente o /dʒ/ em palavras com y, mudando o <i>yes</i> para <i>Jess</i> ou cometem erros com /ʒ/, especialmente porque este som não é muito comum em inglês.</p>
/h/, /f/, /v/	<p>A palavra <i>hello</i> começa com /h/, <i>feel</i> começa com /f/ e <i>very</i> começa com /v/. Tanto o /h/ como o /f/ envolvem o sopro de ar sem a vibração das cordas vocais ou a utilização da língua. O /f/ e o /v/ são muito semelhantes, pois são ambos sons lábio-dentais.</p>

	Existem muitos erros possíveis. Alguns falantes não pronunciam o som /h/ completamente, dizendo <i>eat</i> em vez de <i>heat</i> . Outros, cometem o erro de pronunciar incorretamente o /f/ e /v/. Este tipo de erros pode fazer com que <i>fit</i> soe a <i>pit</i> e <i>very</i> soe a <i>berry</i> .
<i>Nasals</i> (Nasais): /m/, /n/, /ŋ/	As consoantes nasais implicam, como o nome indica, que o som sai pelo nariz e não pela cavidade oral. Existem três sons nasais em inglês. O último som em <i>home</i> é o /m/, o último som em <i>been</i> é o /n/ e o último som em <i>thing</i> é /ŋ/. Muitos falantes cometem erros com estes sons quando estes estão no fim das palavras. Um tipo de erro comum é a troca de sons nasais. Por exemplo, dizer acidentalmente <i>hone</i> em vez de <i>home</i> ou <i>thin</i> em vez de <i>thing</i> .
 /æ/ /ʌ/ /ɑ/	/eɪ/ é a vogal em <i>lack</i> , /ʌ/ é a vogal em <i>luck</i> e /ɑ/ é a vogal em <i>lock</i> . São todas vogais baixas porque o maxilar inferior e a língua estão mais abaixo do que na pronúncia de outras vogais. Um erro comum é não discriminar as três vogais entre si. Por exemplo, muitos falantes cometem o erro de usar o mesmo som da vogal em <i>fan</i> e <i>fun</i> ou em <i>colar</i> e <i>color</i> .
<i>Diphthongs</i> (Ditongos) /eɪ/ /oʊ/ /aɪ/ /ɔɪ/ /aʊ/	Existem cinco ditongos em inglês americano. A vogal e semivogal em <i>say</i> é /eɪ/, a vogal e semivogal em <i>most</i> é /oʊ/, em <i>hi</i> é /aɪ/, em <i>voice</i> é /ɔɪ/ e em <i>how</i> é /aʊ/. Um erro comum é a má pronúncia de um ditongo devido à incorreta colocação da língua. Por exemplo, muitas línguas têm o som /o/ que não envolve o movimento da língua ou lábios. Assim, os falantes dessas línguas podem pronunciar o /oʊ/ americano sem movimento suficiente.
 /u/ /ʊ/	/u/ é a vogal em <i>food</i> e /ʊ/ é a vogal em <i>good</i> . Estas vogais têm algumas semelhanças, mas não são as mesmas. Um erro comum é pronunciar da mesma maneira estas vogais. Por exemplo, alguns falantes dizem <i>pool</i> e <i>pull</i> da mesma maneira, mas, na verdade, estas palavras têm sons diferentes uma da outra.
 /eɪ/ /ɛ/ /æ/	/eɪ/ (<i>fade</i>), /ɛ/ (<i>fed</i>) e /æ/ (<i>fad</i>) têm algo em comum: são todas pronunciadas com a língua na parte da frente da boca, ou seja, são vogais anteriores. Os erros são, normalmente, pronunciar da mesma maneira estas vogais. Por exemplo, alguns falantes combinam acidentalmente /eɪ/ e /ɛ/, dizendo <i>late</i> e <i>let</i> da mesma maneira. Os falantes podem também não pronunciar de forma diferenciada os sons o /ɛ/ e o /æ/ e cometem o erro de dizer <i>set</i> e <i>sat</i> com a mesma vogal.

Tabela 1 – Menus de ajuda da aplicação ELSA Speak



Figura 3 - Sequência de Planetas da aplicação ELSA Speak

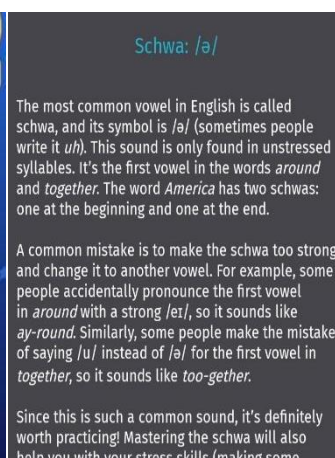


Figura 4 - Menu de ajuda do planeta Schwa da aplicação ELSA Speak

Em todos os planetas o utilizador poderá realizar quatro tipos de exercícios: *pronunciation* (Figura 5), *listening* (Figura 6), *word stress* (Figura 7) e *conversation* (Figura 8). Nos exercícios de *pronunciation* (pronúncia), o utilizador repete a palavra que aparece no ecrã. Nos exercícios de *listening* (audição), o utilizador vai ouvir uma palavra e depois escolher de entre duas palavras similares a que acha que ouviu. Nos exercícios de *word stress* (sílabas tónicas), o utilizador tem de repetir a palavra mostrada no ecrã com o correto acento de sílaba assinalado. Os exercícios de *conversation* (conversa) simulam um diálogo entre o utilizador e a ELSA.

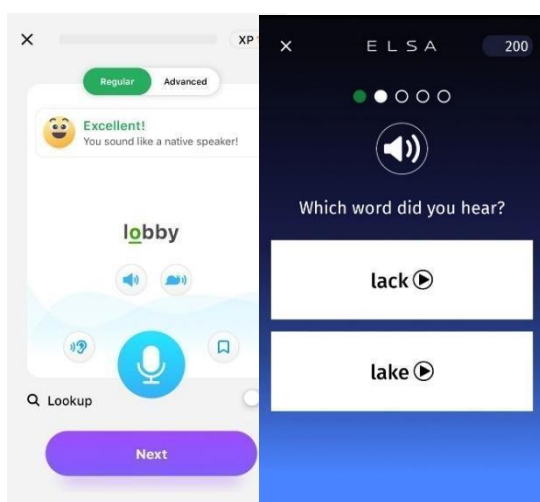


Figura 5 - Exercício de Pronunciation da aplicação ELSA Speak

Figura 6 – Exercício de Listening da aplicação ELSA Speak

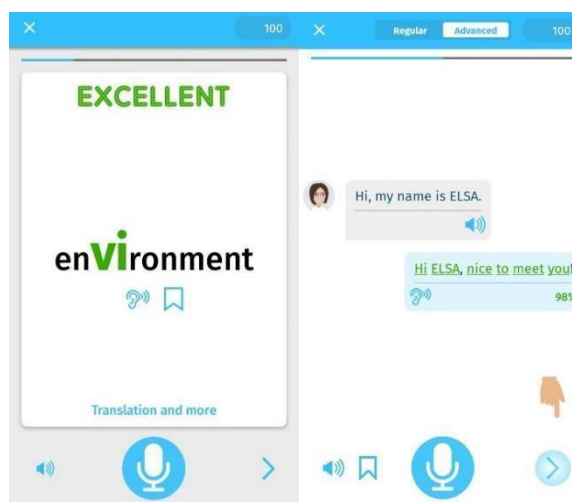


Figura 7 - Exercício de Word Stress da aplicação ELSA Speak

Figura 8 - Exercício de Conversation da aplicação ELSA Speak

A organização das lições da ELSA permite ao utilizador partir de um exercício mais básico para um mais complexo, o que vai promover uma aprendizagem com diferentes níveis de dificuldade. Os utilizadores começam primeiramente por aprender apenas palavras e como estas têm de ser pronunciadas, como por exemplo “Iced Tea”. De seguida, o utilizador vai usar estas mesmas palavras em frases simples para entender o contexto em que estas se apresentam, como por exemplo “I’d like a lemon wedge for my iced tea, please.”, aumentando um pouco o nível de dificuldade. Por fim, nos exercícios de *conversation*, são apresentadas estas palavras em contexto de diálogo, com o intuito

de o utilizador treinar frases ainda mais complexas e, assim, sentir que está a progredir e a imitar contextos reais de diálogo.

A ELSA tem em todos os seus exercícios um relatório fonético com *feedback* (traduzirei *feedback* para comentário, como será doravante referido) de todos os fones pronunciados pelo utilizador. Neste relatório é destacado a verde os fones corretamente pronunciados, a amarelo os fones que não estão corretos, mas são aproximações relativamente à pronúncia alvo, e a vermelho os fones mal pronunciados. Os comentários permitem ao utilizador perceber que sons tem adquiridos e em quais deve ainda investir. Utilizando o IPA, este relatório é criado através de tecnologia de reconhecimento de fala que evidencia os erros dos utilizadores e disponibiliza um comentário, escrito e em áudio, preciso e consistente de como os utilizadores podem corrigir esses erros. Assim, o utilizador vai ter sempre a possibilidade de retificar a sua pronúncia.

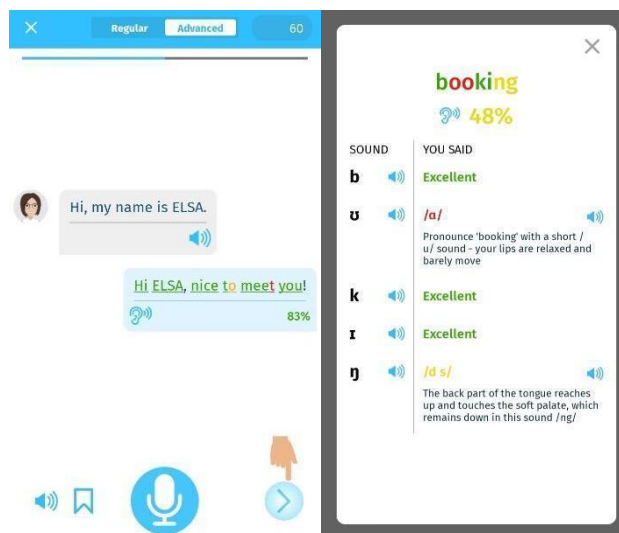


Figura 9 – Comentário fonético breve

Figura 10 – Comentário fonético detalhado

Todos os exercícios têm um modo regular e um modo avançado. É da escolha do utilizador usar um ou o outro. No modo regular, a ELSA oferece um comentário apenas do fone que está a ser testado nesse exercício e no modo avançado o utilizador pode encontrar um comentário em relação a todos os fones da palavra. O modo avançado fornece assim um relatório mais detalhado, permitindo ao utilizador perceber se está a pronunciar corretamente todos os fones.

O utilizador pode também praticar vocabulário específico de um determinado tema. Existem mais de quarenta e cinco tópicos disponíveis no separador *Topics* (Tópicos) e todas as semanas é adicionado novo conteúdo relativo a acontecimentos da atualidade. Os temas destes tópicos vão desde música, feriados, viagens, entrevistas de trabalho, Natal, *Halloween*, entre outros.

Para além dos separadores *Skills* e *Topics*, existe ainda o separador *Dictionary* (Dicionário). Neste separador, o utilizador escolherá uma das quatro palavras apresentadas pela ELSA ou, se preferir, pode escrever uma palavra/frase que pretenda treinar e, de seguida, o utilizador gravar-se-á a pronunciar essa mesma palavra/frase. A ELSA irá depois apresentar-lhe o relatório fonético com o comentário relativo à sua pronúncia.



Figura 11 - Exercício do Dicionário da aplicação ELSA Speak



Figura 12 - Comentário detalhado de um exercício do Dicionário

É importante referir que, para além do comentário fonético, a ELSA também disponibiliza em todos os seus exercícios a definição das palavras do *Oxford Dictionary*, frases com exemplos e a percentagem de falante nativo do utilizador.

O utilizador pode, a qualquer momento, acompanhar o seu progresso e o histórico dos exercícios realizados no separador *Progress* (Progresso). Aqui é possível ver os dias em que praticou e ainda definir um objetivo diário, como por exemplo completar 5 lições por dia. A ELSA enviar-lhe-á notificações para que não se esqueça. Pode também ver o seu nível de proficiência baseado em todos os exercícios que já praticou.



Figura 13 - Separador Progresso na aplicação ELSA Speak

2.3. Tarefas realizadas ao longo do estágio

Após apresentar a companhia e a aplicação, irei agora mencionar quais foram os meus contributos para esta *startup*. Fui estagiária na ELSA Corp. aproximadamente cinco meses e, durante esse tempo, fui responsável por três tarefas. A primeira tarefa foi a tradução para português europeu e português do Brasil da descrição da aplicação na *App Store* e na *Google Play Store*. A segunda tarefa que realizei foi a validação de áudios e correspondentes transcrições em texto e a terceira e última tarefa, na qual me irei focar com mais detalhe ao longo deste relatório de estágio, foi a pós-edição de conteúdo da aplicação traduzido pelos motores de tradução automática Google Tradutor e *DeepL*. Irei, nesta subsecção, descrever em detalhe estas tarefas referindo também qual foi a sua importância para a empresa.

A primeira tarefa na ELSA foi dividida em duas partes. A primeira parte consistiu na tradução da descrição da aplicação ELSA *Speak*, na *App Store*, para português europeu e português do Brasil. Esta descrição é o primeiro impacto que o utilizador vai ter da aplicação e, por isso, é importante que esteja na língua materna (ou na variante dialetal neste caso português europeu e português do Brasil) de potenciais novos utilizadores, de maneira a cativá-los. A segunda parte consistiu na revisão da descrição da aplicação, na *Google Play Store*, já em português europeu e adaptá-la também para português do Brasil. A necessidade de tradução também para português do Brasil deveu-se a um aumento significativo no número de utilizadores brasileiros, que levou também à contratação de

um gestor para a América Latina, que vai analisar todos os conteúdos que serão futuramente adicionados na aplicação. Não sendo falante nativa da variedade de português do Brasil, foi necessário pedir a revisão destes textos a uma falante nativa, de modo a validar o trabalho realizado.

A minha segunda tarefa consistiu na validação de áudios e correspondentes transcrições em texto, para treino e teste do reconhecedor automático de fala. Para este propósito, foram utilizadas as frases mais frequentemente escritas pelo utilizador no separador *Dictionary* (Dicionário), de modo a que o reconhecedor fosse treinado/testado com as frases mais relevantes para os utilizadores e com base em dados ecológicos. Nesta tarefa, tive acesso a essas frases, organizadas em ficheiros Excel, assim como a todas as gravações das mesmas e o meu trabalho consistiu em ouvir as gravações e confirmar que estas coincidiam com o que os utilizadores tinham escrito. Em alguns casos, foram feitas pequenas correções nas frases, para que estas correspondessem integralmente ao áudio, sendo que os erros mais comuns foram os erros ortográficos. A maioria dos utilizadores era de nacionalidade vietnamita, no entanto havia utilizadores de muitas outras nacionalidades. Depois de escolhidas todas as frases relevantes para treinar e testar o reconhecedor de fala, estas foram enviadas a especialistas que desenvolvem o reconhecedor de fala.

A minha tarefa principal no estágio foi a pós-edição de palavras/frases dos diversos exercícios da aplicação, traduzidas automaticamente pelo motor Google Tradutor e *DeepL*, e, assim, assegurar a qualidade de tradução necessária para estas poderem ser corretamente apresentadas aos utilizadores. A aplicação *ELSA Speak*, por ser uma aplicação didática, requer que o utilizador compreenda realmente tudo o que está a ser apresentado nos exercícios. O propósito da *ELSA* não é que os utilizadores apenas imitem palavras/frases, mas que as aprendam de forma a conseguirem reproduzi-las corretamente em situações do dia-a-dia. Assim, a boa qualidade de tradução de todos os exercícios é essencial.

Para realizar esta tarefa, foi-me disponibilizada uma folha Excel com 7059 linhas com palavras e frases dos vários exercícios da aplicação. Este conteúdo estava já traduzido, como foi referido anteriormente, pelos motores de tradução automática Google Tradutor e *DeepL* e o meu trabalho foi a sua pós-edição para português europeu e português do Brasil. Embora com a ressalva de que não sou falante da variedade do português do Brasil,

a empresa considerou que o meu trabalho seria uma boa base para posteriormente um falante nativo poder realizar as edições finais ao conteúdo. Estaria, desta forma, a contribuir para comunicar também com utilizadores brasileiros, como pretendido pela empresa. Assim, visto que a empresa não tem implementado um processo de avaliação de qualidade do texto para português europeu e para português do Brasil, este foi o meu contributo principal.

O trabalho de pós-edição envolveu a adaptação e a correção das diversas frases dos exercícios da ELSA de maneira que estas fossem compreensíveis e adequadas do ponto de vista estilístico ao utilizador. Durante esta tarefa, constatei que os motores de tradução automática utilizados cometem frequentemente alguns erros que podem ser cruciais para uma boa compreensão do conteúdo. Será em alguns dos erros encontrados que me irei focar para a elaboração deste relatório de estágio e que irei explicar e analisar mais detalhadamente na secção 5.

As tarefas que realizei na Elsa foram diversificadas e permitiram-me perceber a dinâmica e a interação entre equipas de linguistas, que fornecem dados de qualidade aos sistemas, e os cientistas de fala, que os processam e tratam com algoritmos neuronais. Permitiu-me também entender todo o trabalho de equipa necessário para garantir a criação e o funcionamento de uma aplicação bem-sucedida como a ELSA *Speak* e ter um pequeno vislumbre do ponto de vista de alguém que desenvolve uma aplicação em vez do de um utilizador comum apenas. Focar-me-ei, como descrito, na análise da tradução de frases do inglês para português europeu e na análise de erros produzidos por motores de tradução automática, nomeadamente Google Tradutor e *DeepL*. Todos os dados utilizados neste relatório foram recolhidos até à data de setembro de 2019. A empresa tem desde esta data desenvolvido a aplicação, criando mais exercícios e versões atualizadas.

3. Estado de arte em Tradução Automática

Neste capítulo será desenvolvido o enquadramento teórico da Tradução Automática para a elaboração deste relatório de estágio. Será primeiramente descrita, na secção 3.1, uma definição de Tradução Automática e o desenvolvimento histórico da mesma, falando de sistemas de tradução baseados em regras e em dados, nomeadamente a tradução estatística e a neuronal. Na secção 3.2, será abordada a Avaliação de Qualidade de Tradução (*Translation Quality Assessment – TQA*) Humana e serão apresentados dois modelos que possibilitaram a sua padronização, LISA (*Localization Industry Standards Association*) QA Model e SAE (*Society of Automotive Engineers*), assim como as suas limitações. Por fim, na secção 3.3, será comentada a tipologia de erros MQM (*Multidimensional Quality Metric*).

3.1. A tradução Automática e o seu desenvolvimento histórico

A Tradução Automática (*Machine Translation – MT*) é a tradução de um texto de um idioma natural (língua de partida) para outro (língua de chegada), feita por um computador. Atualmente, é utilizada por diversas empresas conhecidas, entre elas a Google e a Microsoft, e já faz parte do quotidiano de milhões de utilizadores.

O desenvolvimento da Tradução Automática começa nos anos 30 quando, em 1933, foram apresentadas duas patentes de aparelhos que poderiam ser utilizados como dicionários multilíngues de tradução. Os inventores destes aparelhos foram Georges Artsrouni, em França, e Petr Trojanskij, na Rússia, sendo este último o detentor do aparelho mais avançado (Kenny, 2018: 429).

O que realmente proporcionou o desenvolvimento inicial da Tradução Automática foi, segundo Kenny (2018: 430), um Memorando escrito por Warren Weaver, em 1949. Neste famoso Memorando, Warren lamenta “*the ‘multiplicity of languages’ that impedes cultural interchange between the people of the earth, and is a serious deterrent to international understanding*”, acredita na existência de uma língua universal ainda não descoberta. Warren sugere o uso do contexto das palavras, de forma a eliminar ambiguidades, que se tire proveito do funcionamento lógico da linguagem e o uso de estatísticas e métodos para auxiliar o desenvolvimento da Tradução Automática. Com

este Memorando, Warren proporcionou uma onda de financiamento em pesquisas de Tradução Automática por todo o mundo (Hutchins, 2000: 2).

A 7 de janeiro de 1954 foi feita, pela Universidade de Georgetown, a primeira demonstração pública de Sistemas de Tradução Automática. Esta demonstração, orientada por Léon Dostert em colaboração com a IBM (*International Business Machine Translation*) consistiu na Tradução Automática de 60 frases do russo para inglês e foi realizada para a imprensa que, embora tenha notado algumas falhas no sistema, aceitou com algum entusiasmo as suas virtudes e concluiu que dali a cinco anos haveria sistemas de Tradução Automática capazes de traduzir quase tudo (Hutchins, 2006). O evidente sucesso desta demonstração levou a que nos anos seguintes houvesse um aumento de grupos de pesquisa de Tradução Automática nos Estados Unidos, na União Soviética, no Reino Unido, entre outros países.

Os paradigmas de Tradução Automática utilizados neste período eram os de sistemas baseados nas regras gramaticais e lexicais das línguas, denominados *Rule-based Machine Translation* – RBMT. Estes sistemas utilizam os conhecimentos linguísticos, obtidos através de léxicos e de toda a informação morfológica, sintática e semântica, e aplicam complexas regras para transferir a estrutura gramatical da língua de partida para a da língua de chegada, criando assim texto automaticamente traduzido.

Apesar de ter havido um grande entusiasmo inicial, em 1960 passou a predominar a desilusão. A Tradução Automática não era ainda capaz de ultrapassar a ambiguidade semântica, o que levava a traduções pouco claras. Segundo Yehoshua Bar-Hillel (1960), estas ambiguidades não poderiam ser resolvidas com a ajuda de contexto nem de outras frases semelhantes, mas somente com conhecimento enciclopédico, o que ainda não existia nos computadores da época. Assim, criticou a ideia de *fully high quality translation* concluindo que esta *"is not a reasonable goal, not even for a scientific texts... Reasonable goals are then either fully automatic, low quality translation or partly automatic, high quality translation."* (Bar-Hillel, 1960), visto que para haver qualidade na tradução teria de haver pós-edição humana, de forma a corrigir os erros da Tradução Automática.

Depois desta onda de desilusão, foi criada, em 1964, a ALPAC (*Automatic Language Processing Advisory Committee*), que iria avaliar o desempenho da Tradução Automática dessa época. Este comité, criado por patrocinadores governamentais nos Estados Unidos,

concluiu que a Tradução Automática “*was slower, less accurate and twice as expensive as human translation*” (Hutchins and Somers, 1992: 7) e que “*There is no immediate or predictable prospect of useful machine translation*” (ALPAC, 1966: 32). Depois deste relatório, o financiamento para pesquisa na área da Tradução Automática, que já era pouco, foi ainda mais reduzido.

Apesar do relatório da ALPAC ter levado à diminuição do interesse dos Estados Unidos pela Tradução Automática, o mesmo não aconteceu no resto do mundo. Segundo Kenny (2018: 432), o Canadá e a Europa foram os dois grandes patrocinadores da Tradução Automática nos anos que se seguiram. No Canadá, o inglês e o francês passaram a ser as línguas oficiais o que gerou uma maior procura de traduções neste par de línguas e na Europa, com as Comunidades Europeias (atual União Europeia) havia uma grande demanda de textos em todas as línguas dos Estados-membros.

No Canadá, a Tradução Automática teve mais destaque na área da meteorologia. Foi implementada no sistema Météo, no qual se traduzia relatórios de meteorologia relativamente simples e com frases e vocabulário limitado. Este sistema é, de acordo com Kenny (2018: 432), o exemplo do uso de Tradução Automática mais bem-sucedido do século XX. Este sucesso deveu-se à sua longevidade – em uso de 1977 a 2002 – à sua produtividade – traduzia até 45 mil palavras por dia no início do século XXI – e à qualidade da tradução – nível de precisão de 90%.

Na Europa, a Comissão das Comunidades Europeias (CEC) implementou o sistema de Tradução Automático Systran, em 1976. Este sistema, criado por Peter Toma em 1968, foi considerado um dos mais importantes depois do relatório da ALPAC e inicialmente traduzia apenas para inglês e francês, mas mais tarde incluiu outros pares linguísticos. Foi também utilizado por outras organizações tais como a NATO. Em 2010, o Systran deixou de ser utilizado pela União Europeia, sendo substituído por um novo sistema também criado pela União Europeia, mt@ec. Em novembro de 2017 o sistema foi mais uma vez substituído pela eTranslations (Kenny, 2018: 432).

A década de 80 ficou vastamente marcada pelo aumento de sistemas comerciais, principalmente na América do Norte e na China, projetados para o mercado ainda em desenvolvimento dos computadores pessoais. Estes novos sistemas, *Example-based machine translation* – EBMT, identificavam e extraíam automaticamente de *corpora* paralelo partes do texto e armazenava-as numa base de dados. Quando fosse necessário

traduzir um novo texto, este teria de ser segmentado e o sistema tentava encontrar a tradução correspondente arquivada na sua base de dados (Kenny, 2018: 435). No entanto, Kenny (2018: 432) refere que os textos produzidos por sistemas EBMT precisavam de pós-edição para melhorar a sua qualidade, o que levou a uma grande procura de tradutores. Estes novos sistemas, que seguiam agora um paradigma de Tradução que utilizava *corpora* paralelo para aprender a traduzir, denominados *Corpus-Based Machine Translation* – CBMT, marcaram uma nova abordagem à Tradução Automática e, daí em diante, tornaram-se uma alternativa aos sistemas que até ali se baseavam nas regras linguísticas e lexicais (RBMT).

Em 1988, foi apresentada por um grupo de investigadores da IBM a Tradução Automática Estatística (*Statistical Machine Translation* – SMT). Estes sistemas utilizam *corpora* paralelo de textos bilíngues do Parlamento Canadano, para criar modelos de probabilidade de tradução que conseguiam produzir boas traduções sem a utilização de regras linguísticas (Kenny, 2018: 433).

Apesar de serem inicialmente recebidos com alguma descrença, os sistemas de SMT foram amplamente desenvolvidos no início dos anos 2000. Segundo Kenny (2018: 433), a disponibilidade de bitextos eletrónicos, o aumento do poder computacional e do armazenamento de dados e o esforço colaborativo de desenvolvedores que disponibilizaram ferramentas *open-source* de SMT proporcionaram este desenvolvimento. Outro agente importante para o desenvolvimento de sistemas SMT foi o fator económico e o fator geopolítico. Em 2000, os Estados Unidos mostraram interesse na Tradução Automática para mandarim e, em 2001, com o ataque às Torres Gémeas, financiaram também pesquisa de Tradução Automática para árabe. Na Europa, com o crescimento da União Europeia, que tinha em 2004, 20 línguas oficiais e em 2014, 24, estava a aumentar a sua necessidade de traduzir os textos nas várias línguas dos Estados-membros (Kenny, 2018: 433).

O aumento do acesso à internet permitiu que entre os anos 90 e início de 2000 a Tradução Automática pudesse crescer e ser utilizada por milhares de utilizadores. Segundo Kenny (2018: 433), havia agora uma diversidade linguística *online* devido às inúmeras páginas *web* multilingues disponíveis, dados que serviam de base para que a SMT se desenvolvesse. Em 1997 a Systran estava a ser usada no motor de pesquisa Alta Vista e em 2006 foi criado o Google Tradutor, que foi, mais tarde, integrado em diversas

aplicações. Começou também a existir uma interação com os utilizadores que agora podiam enviar comentários das traduções. A Microsoft foi também outra grande empresa tecnológica a criar sistemas de Tradução Automática.

Apesar da evidente aceitação da Tradução Automática por parte dos utilizadores *online*, esta não era ou é ainda bem vista pelos tradutores profissionais. Segundo Kenny (2018: 433), a razão para tal foi não só por o resultado da Tradução Automática ser ainda de má qualidade, não valendo mesmo a pena a tarefa de pós-edição, por ser necessário traduzir de raiz, mas também por terem aparecido no mercado as Memórias de Tradução (*Translation Memories*), que permitem a reutilização de traduções feita por profissionais. As Memórias de Tradução aumentaram a produtividade dos tradutores e permitiram o crescimento da automatização da tradução, pois possibilitaram, por um lado a criação de bitextos em grandes quantidades, que podiam ser lidos por sistemas SMT e, por outro, a pós-edição por humanos com a ajuda de Memórias de Tradução já feitas já realizadas anteriormente.

Os sistemas SMT foram difundidos por todo o mundo e foram considerados os mais avançados e eficientes. No entanto, em 2015/16, foi desenvolvida a Tradução Automática Neuronal (*Neural Machine Translation – NMT*), um sistema promissor e com um melhor desempenho, que, tal como os sistemas SMT, aprende a traduzir através de *corpora* paralelos, no entanto, utiliza modelos de redes neuronais artificiais (semelhantes às do cérebro humano), para desenvolver os modelos estatísticos. Assim, os sistemas SMT foram substituídos por sistemas NMT e, no fim de 2016, estes estavam já integrados em diversas empresas e aplicações tais como a Google e a Microsoft. (Kenny, 2018: 434).

Ao longo do século XX, a Tradução Automática teve um enorme desenvolvimento e passou por variadas alterações, promovidas principalmente pelo avanço tecnológico. É possível constatar que existem dois grupos de paradigmas principais de Tradução Automática: Os paradigmas dos sistemas que seguem as regras linguísticas e lexicais, intitulados *Rule-based Machine Translation – RBMT*, e os paradigmas dos sistemas que utilizam *corpora* paralelo para comparar textos já traduzidos e, deste modo, calcularem a tradução mais provável, denominados *Corpus-based Machine Translation – CBMT*, e no qual estão integrados os sistemas EBMT, SMT e NMT.

Atualmente, os sistemas NMT substituíram os sistemas SMT e são considerados pelos especialistas os mais complexos e proeminentes. Com o intuito de identificar e analisar

em que aspetos são os sistemas NMT superiores, Bentivogli *et al.* conduziu um estudo onde comparou os resultados de tradução de sistemas NMT com os de sistemas SMT. Neste estudo, Bentivogli *et al.* (2016: 9) refere que “*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*”. Em relação à ordem das palavras, menciona que “*NMT demonstrated a 70% reduction in the incorrect placement of verbs*” e ainda que os esforços de pós-edição eram consideravelmente mais baixos (-26%), concluindo que sistemas NMT são consideravelmente mais avançados do que sistemas SMT (2016: 9). Por outro lado, Castilho *et al.* (2017, 118) discorda e refere que “*While automatic evaluation results published for NMT are undeniably exciting, so far it would appear that NMT has not reached the quality of SMT, based on human evaluation*”. Apesar de serem bem aceites entre os especialistas, os sistemas de NMT mostram ainda alguns problemas nas suas traduções finais. Bentivogli *et al.* (2016: 9) menciona que sistemas NMT têm ainda algumas limitações que precisam ser ultrapassadas, tais como “*handling of long sentences and reordering of particular linguistic constituents requiring a deep semantic understanding*”. Já Castilho *et al.* (2019: 2) refere que a ascensão dos sistemas NMT levou também à crescente consciencialização de que é fundamental ultrapassar essas limitações e assegurar uma melhor qualidade das traduções. Para tal, Castilho propõe a investigação mais aprofundada de amostras maiores, a intervenção de revisores (tradutores profissionais, alunos de tradução e utilizadores com diferentes níveis de experiência) na pré-edição e pós-edição das traduções e para mais pares linguísticos (2019: 5).

3.2. A Avaliação de Qualidade de Tradução e primeiros esforços de padronização

A tradução passou a ser de grande importância e é utilizada em diversas áreas e empresas de grande prestígio. Assim, garantir a qualidade dos textos traduzidos é essencial. A Avaliação de Qualidade de Tradução (*Translation Quality Assessment – TQA*, como será doravante referida) é um tópico assaz importante na área da tradução. Dada a sua importância, foram feitos vários esforços por parte de especialistas para chegarem a um consenso sobre o que é TQA e como esta deve ser definida e analisada, porém, os

especialistas afirmam que não existe apenas uma maneira de aferir a qualidade de uma tradução. O objetivo da TQA é avaliar a qualidade da tradução, qualidade definida segundo especificações impostas pelo cliente. Para isso, o revisor das traduções vai identificar os tipos de erros que possam existir, contabilizá-los e atribuir-lhes uma pontuação. No entanto, avaliar a qualidade de uma tradução é algo complexo e foi um tema que durante algum tempo causou discórdia entre os especialistas e ainda persiste.

A inexistência de um modelo padrão de TQA contribuiu para a divergência de opiniões entre os especialistas em relação à definição do que realmente a TQA comporta. Uma das razões para este desentendimento é o facto de “qualidade” ser um termo que pode variar de pessoa para pessoa, ou neste caso, de cliente para cliente. Inicialmente, cada revisor (pessoa que vai avaliar a qualidade da tradução) avaliava as traduções de acordo com as suas próprias normas, dando maior ênfase a certo tipo de erros e desconsiderando outros, no entanto, segundo Lommel (2018: 110), este método deixava ainda muito a desejar, pois cria uma enorme discordância na avaliação das traduções e na identificação e classificação dos tipos de erros.

Com o objetivo de padronizar a TQA, foram desenvolvidos, em 1990, dois modelos de avaliação de tradução humana que merecem ser destacados. De acordo com Lommel (2018: 112), o primeiro modelo foi o da SAE (*Society of Automotive Engineers*) J2450, usado para a avaliação da qualidade de documentos relacionados com automóveis, e o LISA (*Localization Industry Standards Association*) *QA Model*, que continuou a ser utilizado em diversas atualizações, mesmo depois de a LISA ter cessado as suas operações em 2011. Estes foram os modelos considerados padrão na altura e os que serviram de exemplo para, mais tarde, o desenvolvimento dos modelos dos dias de hoje.

Nestes modelos, as traduções eram avaliadas a partir de um sistema de pontuação. Os revisores utilizavam listas de tipos de erros e atribuíam a cada segmento do texto uma pontuação, consoante o número de erros aí existentes. O modelo SAE considerava 7 tipos de erros: termo errado (*wrong term*), erro sintático (*syntactic error*), omissão (*omission*), erro na estrutura ou concordância das palavras (*word structure or agreement error*), erro ortográfico (*misspelling*), erro de pontuação (*punctuation error*) e erros variados (*miscellaneous error*). Já o LISA *QA Model* era um modelo mais abrangente e avaliava entre 18 e 21 tipos de erros, representados na seguinte tabela:

Documentation	Help	Software
Doc Language (7) <ul style="list-style-type: none"> • Mistranslation • Accuracy • Terminology • Language • Style • Country • Consistency Doc Formatting (6) <ul style="list-style-type: none"> • TOC • Index • Layout • Typography • Graphics • Call Outs and Captions Doc Formatting – Asian (8) <ul style="list-style-type: none"> TOC Index Layout Typography Graphics Call Outs and Captions Double/Single Size Punctuation Marks 	Help Formatting (4) <ul style="list-style-type: none"> • Index • Layout • Typography • Graphics Help Formatting – Asian (4) <ul style="list-style-type: none"> • Index • (Layout) • Graphics • (Typography) • Localizable Text • Hyper text • functionality, jumps, popups 	Software Formatting (5) <ul style="list-style-type: none"> • Graphics • Alignment • Sizing • Truncation/Overlap • Character Formatting Software Functionality Testing (5) <ul style="list-style-type: none"> • Localizable Text • Dialogue Functionality • Menu Functionality • Hotkeys/Accelerators • Jumps/Links
21 total	8 total	10 total
TOTAL: 39 total (25 unique)		

Figura 14 - Tipos de erro do LISA QA Model

Cada erro podia ser classificado com um grau de gravidade (*minor*, *major* ou *severe*) e iria penalizar a pontuação final da qualidade. A tradução podia “passar” ou “falhar”, consoante o limite máximo de erros imposto pelo cliente. A pontuação final era em percentagem, sendo que 100% significava que não existiam erros.

No entanto, estes modelos de avaliação mostravam ter ainda algumas limitações. Apesar de já permitirem a troca de resultados entre revisores, algo inédito até ali, mantinha-se o problema da ambiguidade em relação à identificação de erros e do seu grau de gravidade. De acordo com Lommel (2018: 112), tal como acontecia nos primeiros modelos de avaliação, no SAE e no LISA QA Model, o que poderia ser um erro para um revisor poderia não o ser para outro, ou podia não ter a mesma importância e, por isso, o mesmo grau de gravidade. Outra limitação, principalmente do LISA QA Model, foi que ao tentar criar um padrão de erros que podia ser utilizado em determinados textos/contextos, estes poderiam não ser apropriados para outros. O SAE, apesar de alegar ser um modelo para o contexto específico da área dos automóveis, era utilizado também para avaliar textos de outras áreas (Lommel, 2018: 113). Uma outra limitação foi o facto de serem modelos “one-size-fits-all”, o que dificultou a sua adaptabilidade às especificações individuais de tradução requeridas pelo cliente para a sua avaliação (Castilho *et al.*, 2018: 15).

3.3. A tipologia de erros MQM

A *Multidimensional Quality Metric* – MQM – é o mais recente esforço de uniformização da TQA. Esta métrica foi desenvolvida pelo projeto QTLauchPad, financiado pela União Europeia, e é, segundo Castilho *et al.* (2018: 16), baseado em métricas partilhadas, que pode ser utilizado tanto na avaliação da Tradução Humana, como da Tradução Automática e em qualquer género de texto, o que possibilita a comparação dos diferentes resultados das avaliações de qualidade. A definição de MQM é, tal como estabelecida pelos autores: “*a framework for describing and defining custom translation quality metrics. It provides a flexible vocabulary of quality issue types, a mechanism for applying them to generate quality scores, and mappings to other metrics. It does not impose a single metric for all uses, but rather provides a comprehensive catalog of quality issue types, with standardized names and definitions, that can be used to describe particular metrics for specific tasks*” (<http://www.qt21.eu/quality-metrics/>). Este modelo utiliza listas de erros, comuns a todos os revisores, para descrever e classificar tipos de erros e aplica métricas de qualidade que unem a qualidade de tradução às especificações do utilizador (Lommel, 2018: 109) e tem agora a flexibilidade necessária para acomodar outros critérios de avaliação que modelos anteriores não podiam (Castilho *et al.*, 2018: 17).

Com o rápido crescimento e a necessidade de tradução de textos para as 27 línguas oficiais, a União Europeia precisava de melhorar as suas capacidades de Tradução Automática. Um estudo da META-NET Language White Papers mostrou que havia ainda muitas línguas oficiais que careciam do apoio necessário de Tradução Automática. Muitas destas línguas são morfologicamente complexas e não tinham ainda disponíveis as ferramentas necessárias/adequadas para a sua tradução, o que levava a traduções de má qualidade. Esta barreira linguística impedia a livre circulação de pessoas, informação e comércio no mercado único digital europeu. Assim, foi criado o projeto Quality Translation 21 - QT21, desenvolvido pelo programa de Investigação da União Europeia, Horizon 2020, que pretende melhorar a qualidade das Tradução Automática (<http://www.qt21.eu/?target=Introduction>), e do qual foi desenvolvida a MQM do QT21. Em 2014, a MQM harmonizou-se com a DQF (*Dynamic Quality Framework*), o modelo da TAUS (*Translation Automation User Society*), de modo a solucionar as preocupações

dos especialistas que referiam que o desenvolvimento em separado da MQM e da DQF estava a atrasar a adoção das melhores práticas de qualidade.

A harmonização da DQF e da MQM veio favorecer a abordagem à TQA. Foram adicionadas à sua estrutura várias categorias principais permitindo, assim, analisar até 50 tipos de erros. A MQM/DQF, (MQM, como será doravante referida) apresenta uma estrutura hierárquica representada na seguinte figura:

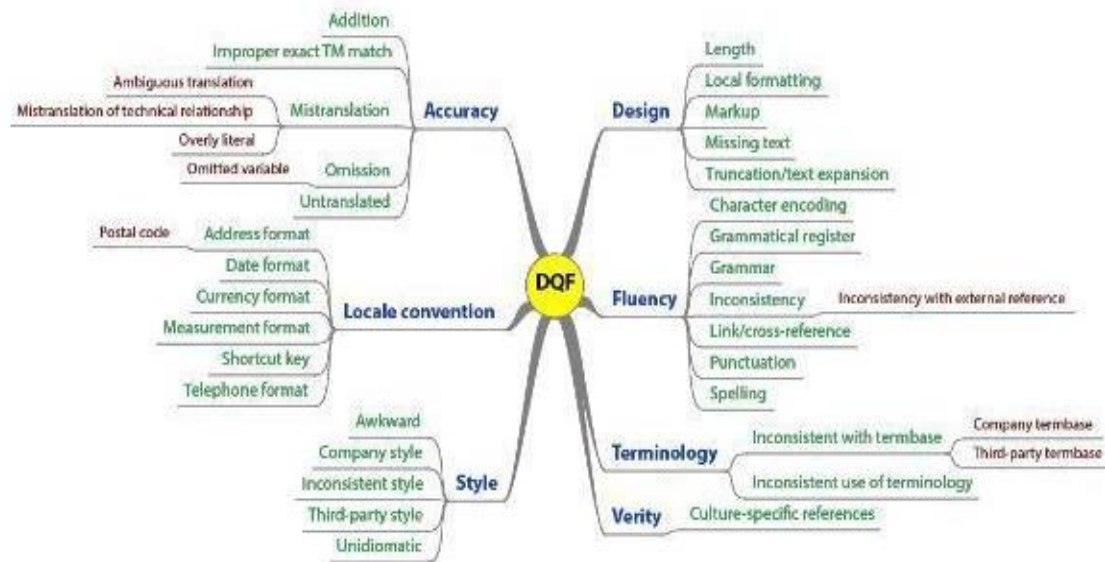


Figura 15 - Estrutura Hierárquica da MQM/DQF

Existem 7 categorias de erros principais e cada uma destas vai avaliar uma determinada característica do texto:

- *Accuracy* - Relação entre o texto de partida e o texto de chegada. Existe um erro quando o conteúdo proposicional do texto de chegada não corresponde ao conteúdo do texto de partida.
- *Fluency* - Questões relacionadas com a boa formação linguística do conteúdo, como por exemplo, se existem erros que não permitem que o texto seja compreendido.
- *Design* - Questões relacionadas com o aspeto gráfico do texto, como por exemplo, a formatação.
- *Locale convention* - Questões que abordam se o conteúdo está a ser corretamente apresentado para a língua e cultura de destino. Por exemplo, as datas são exibidas em diferentes formatos, consoante o país;

- *Style* – Questões relacionadas com problemas estilísticos no texto, como por exemplo, a tradução de uma campanha publicitária leve e humorística para um estilo mais sério, embora as especificações dissessem que deveria corresponder ao estilo do texto original.
- *Terminology* - Questões relacionadas com a utilização de terminologia específica do domínio do conteúdo do texto;
- *Verity* – Questões relacionadas com o conteúdo do texto de chegada que não está de acordo com o seu verdadeiro contexto, como por exemplo, se um manual de instruções afirmar que um automóvel tem uma determinada característica que, na realidade, não tem.

De modo a simplificar a sua utilização, foi também elaborada a MQM Core, que contém os 20 tipos de erros mais comuns nas traduções. A MQM Core é encorajada pelos especialistas com o propósito de promover uma maior interoperabilidade entre os resultados obtidos nos diferentes textos. O utilizador pode utilizar as categorias que pretender, pode aplicar métricas mais simples ou métricas mais detalhadas que podem incluir várias categorias. No entanto, é recomendado que, caso faça uma avaliação de qualidade de tradução mais simples, utilize pelo menos as categorias fundamentais *Accuracy* e *Fluency*. (<http://www.qt21.eu/mqm-definition/definition-2015-06-16.html#mqm-core>).

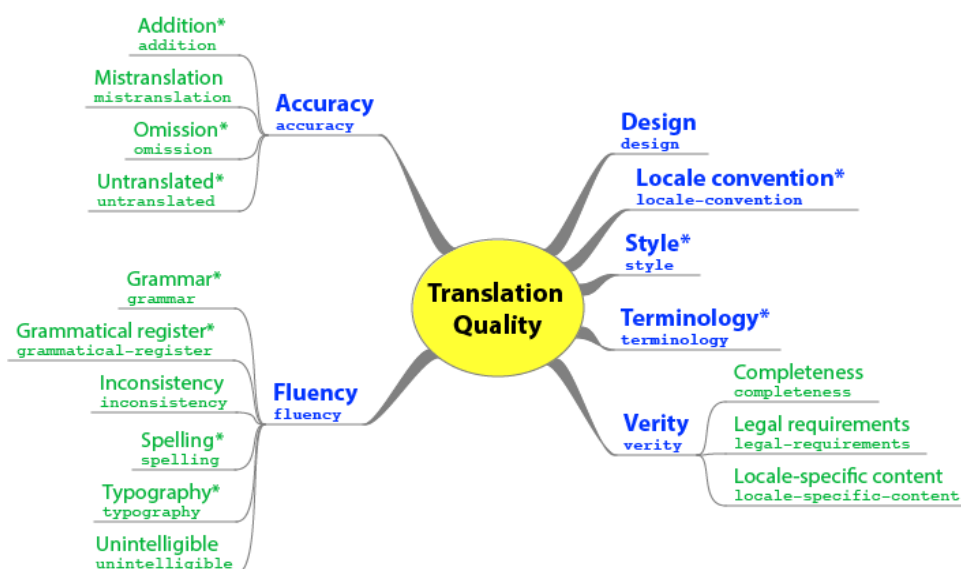


Figura 16 – MQM Core

Outro aspeto que é importante referir sobre a MQM são os níveis de severidade atribuídos aos erros encontrados nos textos. A MQM tem disponíveis quatro níveis de severidade (*none*, *minor*, *major* e *critical*) e cada um destes níveis vai ter um peso na pontuação final. Um erro “none” serve como uma chamada de atenção e não tem qualquer impacto na pontuação final, um exemplo é o uso constante de um termo de domínio incorreto que pode ser facilmente corrigido. Erros “minor” (menores) não têm impacto na usabilidade do conteúdo do texto nem na sua compreensão, como por exemplo, um espaço extra depois de um ponto final. Os erros “major” (graves) têm impacto na usabilidade ou na compreensibilidade do conteúdo do texto, mas não o tornam inutilizável, como por exemplo, uma palavra mal escrita. Já os erros “critical” (críticos) são erros que não permitem a utilização do conteúdo do texto, como por exemplo, um erro gramatical que muda o significado do texto. (<http://www.qt21.eu/mqm-definition/definition-2015-06-16.html#scoring>)

Depois de anotados todos os erros do texto, é aplicada a fórmula da métrica MQM (AMTA, 2016: 21), com o intuito de calcular a percentagem de qualidade de tradução. Esta fórmula é calculada através do número total de palavras do texto e do número de erros *critical* (críticos), *major* (graves) e *minor* (menores) existentes e está representada da seguinte maneira:

$$\text{MQM Score (\%)} = 100 - ((\text{IssuesMinor} + 5 * \text{IssuesMajor} + 10 * \text{IssuesCritical}) / \text{Sentence length}) * 100.$$

Textos com um resultado de 80% são considerados pouco satisfatórios e vão precisar de edição humana. Textos com 90% têm uma boa qualidade e textos com resultados iguais ou superiores a 95% têm uma qualidade excelente, sendo equiparados a uma tradução humana.

A MQM foi desenvolvida de forma a evitar algumas limitações encontradas nos modelos de avaliação de qualidade anteriormente apresentados e, por isso, foi tão bem aceite entre os especialistas. De acordo com Lommel (2018: 114), o que contribui para o seu sucesso foi o facto de este modelo adotar quatro princípios fundamentais:

i) Dispõe de uma lista flexível de tipos de erros. Disponibiliza um vasto vocabulário para definir e classificar as diferentes categorias de tipos de erros de tradução. Assim, os

revisores conseguem facilmente determinar a que categoria se aplica um determinado erro.

ii) Foi desenvolvido com base nas especificações e ferramentas já existentes, sendo assim compatível com estas ferramentas, o que facilitou a sua difusão;

iii) Adota uma estrutura hierárquica. Uma vez que nem todos os textos necessitam de uma abordagem muito detalhada, foi necessária a criação de uma estrutura de avaliação que permitisse incluir categorias de erros mais gerais e outras mais específicas. A MQM tem uma estrutura em árvore, composta inicialmente por 8 ramos principais (categorias de erros gerais) que se vão segmentar em outros ramos mais pequenos (erros mais específicos);

iv) Tem uma abordagem baseada na especificação dos erros a serem analisados. Devido a esta abordagem é possível que, no caso de um tipo de erro que não seja um requisito do cliente para a avaliação da qualidade da tradução, este erro possa ser desconsiderado. Assim, a pontuação final da tradução não vai ser injustamente penalizada. Por exemplo, se o *Style* de uma tradução não for um requisito, o revisor não vai poder penalizar a pontuação, caso haja efetivamente problemas deste tipo.

De acordo com Castilho *et al.* (2018: 17), foram feitos vários esforços para alcançar modelos eficazes de avaliação da qualidade que combinam, ou até reconciliam, diferentes pontos de vista sobre o que é a qualidade da tradução. A uniformização da TQA, resultado do desenvolvimento da MQM, permitiu, segundo Lommel (2018: 109), escapar à inconsistência e à subjetividade que até ao momento da sua utilização alargada caracterizavam a TQA.

3.4. Diálogos Informais

Em diálogos, especialmente em diálogos de registo informal e espontâneo, é frequente encontrar diversos marcadores conversacionais e outros mecanismos linguísticos. Estas características, tipicamente “orais”, são produzidas de forma natural e frequente pelo falante nativo. Para compreender verdadeiramente uma língua, é necessário ter não só os conhecimentos gramaticais dessa mesma língua (utilizados em textos escritos), mas também saber como esta pode ser utilizada em diversos contextos sociais (em diálogos)

e situacionais. Segundo o Dicionário Terminológico da Direção Geral da Educação, o discurso oral prototípico “tem uma sintaxe pouco estruturada² – com predomínio, como ficou dito, da parataxe, com orações incompletas, com repetição de estruturas, com elipses, etc. – e, em geral, apresenta uma formalidade mais débil e descuidada do que o texto escrito.” (<http://dt.dge.mec.pt/index.php?id=n442>). No entanto, o discurso pode também ser bem estruturado e seguir as regras gramaticais noutras situações, como por exemplo, nos discursos políticos, que são habitualmente mais formais. A livre tomada de turnos é também uma característica representativa dos diálogos, sendo que cada tomada de turno representa uma intervenção do interlocutor.

Segundo a Gramática do português (Volume III: cap. 51), num cenário de discurso “o locutor recorre naturalmente ao seu conhecimento da língua, um saber gramatical, mas mobiliza igualmente outros saberes, de natureza sócio-comunicativa e cultural, que relevam basicamente de a necessidade de adequar o discurso ao assunto de que se fala (o tópico), aos objetivos da interação, ao perfil do interlocutor, à especificidade da própria situação de comunicação.” A adequação do discurso ao contexto pragmático vai influenciar o tipo de registo utilizado pelo locutor, apropriado a esse mesmo contexto. Existem dois tipos de registo: formal e informal. A utilização de um ou de outro registo depende do grau de formalidade que os interlocutores têm entre si, do tópico e até do local onde ocorre o discurso (no Parlamento Europeu é esperado um discurso mais formal do que, por exemplo, numa conversa entre dois amigos). Nuno Santos (2019: 14) corrobora para a tradução de diálogos e adaptações realizadas e refere que a utilização de um ou outro registo é o resultado da adaptação do discurso, por parte do falante, consoante a situação e o contexto em que este se encontra, levando-o a interagir de acordo com as regras. Assim, o contexto é o fator determinante que leva o locutor a utilizar ou o registo formal ou o informal e vai ter influência na construção sintática das frases, nas escolhas lexicais, nas formas de tratamento entre os interlocutores, entre outros (Gramática do português, Volume III: cap. 51).

Os diálogos informais são naturalmente espontâneos e não planificados, no entanto, segundo Garcia (2009: 34), isto não significa que careçam de uma estrutura ou que esta seja demasiado variável ou caótica para ser analisada. Na verdade, ainda segundo Garcia

² É discutível que a sintaxe da oralidade seja realmente pouco estruturada. Para uma discussão crítica, veja-se Moniz 2013 (tese de doutoramento).

(2009: 34) os diálogos têm “*mecanismos principios bien definidos y está perfectamente organizada a nível global (la estrutura geral) Y a nivel local (los turnos y sus relaciones*”, sendo que o nível global corresponde ao macro nível e o nível local ao micro nível, de acordo com Van Dijk (1989: 261). Assim, considerando a estrutura global do diálogo, Moreno (2002: 55) identifica as fases prototípicas da conversação, representadas na seguinte tabela:

Fase da conversação	Objetivo	Exemplo
Preparação	Chamar a atenção para iniciar o diálogo.	Através de sinais verbais (“Olhe/Olha”; “Escuta”) ou não-verbais (Aceno de cabeça; Sorriso)
Abertura	Saudação entre os interlocutores que depende da relação de proximidade entre eles.	“Olá”; “Bom dia”;
Preâmbulo	Várias perguntas de rotina sobre a saúde do interlocutor e a da sua família, o trabalho, etc.	“Como está/estás?” “Como vai o trabalho?”
Orientação	Serve para direcionar o diálogo para o objetivo/assunto da conversa, através de perguntas ou comentários sobre algo.	“Já sabes o que se passou ontem?” “Nem imaginas o que aconteceu?” (Van Dijk, 1989:278)
Objetivo do diálogo	Assunto do diálogo	Temas diversificados
Conclusão	Reconhecimento de ambas as partes que o diálogo está concluído.	Expressões como “enfim”, “pois” ou “pronto”
Encerramento	Despedida dos interlocutores que depende da relação de proximidade entre eles.	Através de formais verbais (“Adeus”) ou não-verbais (Aperto de mão, abraço ou dois beijos)

Tabela 2 – Fases prototípicas da conversação segundo Moreno (2002)

As fases da orientação, do objetivo do diálogo e da conclusão são categorias recursivas, ou seja, podem ser repetidas diversas vezes durante o diálogo.

De acordo com Garcia (2002: 36), a fase da abertura e a da conclusão são influenciadas pelo contexto sociocultural em que ocorre o diálogo e pela relação de afinidade entre os

interlocutores. Por este motivo, é possível encontrar “*un gran número de fórmulas ritualizadas, tanto verbales como no verbales, particulares de cada cultura*” (Garcia, 2009: 36).

Apresentada uma breve resenha da estrutura global da conversação e das fases que a compõem, será agora analisada a microestrutura da conversação. Neste nível, serão ponderadas tomadas de palavra e as suas relações. Uma tomada de palavra equivale a uma intervenção do interlocutor no diálogo e, seguindo a proposta teórica de Sacks, Schegloff & Jefferson (1974), cada tomada de palavra simboliza uma unidade básica da estrutura interna da conversação e a sucessiva alternância de tomadas de palavra é o que vai promover e sustentar todo o processo conversacional entre os diversos interlocutores. Assim, de acordo com Pimentel (2012: 20), pode concluir-se que “essa alternância de turnos origina pares conversacionais ou pares adjacentes, como por exemplo, pergunta-resposta, convite-aceitação/recusa. Uma sucessão de turnos constitui uma sequência e várias sequências constituem uma conversação.” Segundo Marcuschi (2003: 18), os silêncios e os sinais verbais e não-verbais dos interlocutores podem também ser considerados uma tomada de palavra ou turno.

Num diálogo, existem variadas estratégias conversacionais que têm como objetivo manter a conversa fluída e coerente. Segundo Garcia (2009: 17), estas estratégias passam por “*hacer contribuciones relevantes y coordinadas, asegurarse la atención del interlocutor, tomar y ceder la palabra de forma apropiada y sincronizada, controlar las pausas y la velocidad de habla, introducir temas a través de los procedimientos adecuados y saber susperderlos y retomarlos cuando la ocasión así lo exija; en definitiva, todo un conocimiento de las estructuras, esquemas, normas y expectativas propios de una comunidad.*” Saber em que fases da conversação se devem utilizar as estratégias acima referidas é também crucial para o bom diálogo entre os interlocutores. De acordo com Martin (2000: 1), um bom conversador sabe:

- Iniciar a conversa de forma adequada, saudando se for necessário, captando a atenção do interlocutor, mostrando disponibilidade de querer participar no diálogo;
- Manter o diálogo seguindo um conjunto de estratégias:
 - Controlar as mudanças de turnos, as pausas e os silêncios;

- Controlar e resolver os mal-entendidos, fazendo perguntas ou reformulando as frases;
- Transmitir sinais de que está a ouvir, mostrando que está atento e a incentivar o outro interlocutor a continuar a conversa;
- Contribuir espontaneamente com informações, opiniões ou conjeturas;
- Mostrar se está de acordo ou de desacordo;
- Intercalar espontaneamente com piadas.

– Finalizar de forma ordeira e não bruscamente, dando sinais de que a conversa está a terminar.

Estas estratégias de conversação são fundamentais para a dinâmica entre os interlocutores e são implementadas no diálogo através de, entre outras estruturas linguísticas, marcadores conversacionais. Segundo a Gramática do português, a utilidade dos marcadores conversacionais é:

– Assinalarem a tomada de turno de fala (*Bem, bom, então, então é assim, ora bem, pois, ...*) ou a cedência de turno de fala (*enfim, não é, pronto, ...*);

– Marcar o processo de formulação não planificada do discurso, assinalando, por exemplo, momentos de hesitação (*digamos, eh, portanto, quer dizer, ...*) ou de autocorreção (*ou melhor, ou seja, ...*);

– Função fática que corresponde à tentativa de manter vivo o contacto com o interlocutor (*olha, percebes, tás a ver, ...*) e à sinalização de atitude de atenção cooperante por parte do interlocutor (*certo, claro, pois, sim, ...*).

A função destes marcadores é “estabelecer a conexão entre enunciados, organizando-os em blocos, indicando o seu sentido argumentativo, introduzindo novos temas, mantendo e orientando o contacto do locutor com o interlocutor” (<http://dt.dge.mec.pt/index.php?id=n442>), isto é, os marcadores são essenciais para garantir a eficiente comunicação entre os interlocutores uma vez que vão facilitar um diálogo fluente e interativo.

Assim, os diálogos informais, apesar de espontâneos, seguem uma aparente estrutura que pode ser dividida em sete fases, cada uma com uma finalidade específica. No decorrer destas fases, os interlocutores passam pelo processo de tomada de palavra ou turno, sendo que uma sequência de tomada de palavra vai estabelecer o diálogo. Com o objetivo de

que o diálogo seja coerente e fluido é também necessário que os interlocutores sigam algumas estratégias de conversação que passam, por exemplo, por controlar as mudanças de turnos, as pausas e os silêncios, transmitir sinais de que está a ouvir, mostrando que está atento e a incentivar o outro interlocutor a continuar a conversa, entre outras. Estas estratégias são implementadas por marcadores conversacionais.

4. Metodologia e Corpus

Os objetivos deste relatório de estágio são a análise de variados diálogos, traduzidos de inglês para português pelos motores de tradução automática Google Tradutor e *DeepL*, e a análise dos erros mais frequentemente produzidos por estes mesmos motores. O intuito deste estudo é compreender se os sistemas de tradução automática dos dias de hoje, utilizados diariamente por milhares de utilizadores *online*, estão já preparados para uma tradução de qualidade de diálogos informais e, desta forma, aferir da qualidade final das traduções produzidas.

Os exercícios de conversação da aplicação *ELSA Speak* são de grande importância, pois são nestes exercícios que o utilizador vai conseguir aplicar palavras praticadas noutros exercícios, em contextos específicos, e treinar frases em contextos pragmáticos distintos. No entanto, a finalidade destes exercícios não é que os utilizadores apenas repitam frases, é necessário que estes compreendam o que as frases significam na sua língua nativa para, assim, as poderem usar adequadamente no quotidiano. Para tal, é necessário garantir uma boa tradução das mesmas.

Neste capítulo, será primeiramente apresentado, na secção 4.1, o corpus que me foi disponibilizado pela empresa e como este está dividido. Na secção 4.2, será apresentada a base de dados que contém todos os erros que são objeto de estudo para a realização deste relatório. Na secção 4.3, serão explicados quais os critérios de anotação de erros que foram seguidos para a criação da base de dados.

4.1. Corpus

Para o cumprimento da minha tarefa principal na *ELSA Corp.*, foi-me facultado um corpus com conteúdos utilizados nos exercícios da aplicação *ELSA Speak*. Este conteúdo pode variar entre apenas palavras ou frases completas e está agrupado consoante o módulo e o tipo de exercício. O corpus está em formato Excel, contém 7130 linhas e está dividido em dez colunas, sendo que cada uma delas corresponde a que a seguir se enumera:

- `Module_id`: a identificação do módulo a que a palavra ou frase pertence;
- `Lesson_id`: a identificação da lição a que a palavra ou frase pertence;

- Game_type: o tipo de jogo (*pronunciation, listening, word stress e conversation*);
- Exercise_id: a identificação do exercício;
- Sentence: a frase ou palavra a ser traduzida;
- Google Tradutor: a tradução sugerida pelo Google Tradutor;
- DeepL: a tradução sugerida pelo DeepL;
- PT: a proposta de tradução em português europeu;
- PT – BR: a proposta de tradução em português do Brasil.

A tarefa principal na empresa envolveu, por um lado, analisar as traduções do Google Tradutor e do *DeepL* de todos os exercícios da aplicação, e, por outro, pós-editar as sugestões dos sistemas, de forma a alcançar um bom resultado final e, assim, evitar que sejam mostrados exercícios com erros de tradução ao utilizador.

4.2. Construção da base de dados para a análise de erros de tradução

Depois de concluída a revisão e a pós-edição do conteúdo da aplicação, foi primeiramente criada uma base de dados que engloba os erros cometidos pelo Google Tradutor e pelo *DeepL* e a atribuição da tipologia de erro e grau de severidade a cada exemplo. Para a elaboração desta base de dados, foram selecionados, do corpus da ELSA, apenas os exercícios de conversação que serão o foco deste relatório. Assim, foram considerados 186 diálogos. A base de dados foi construída em formato Excel e está distribuída da seguinte maneira:

- Frase original;
- Tradução produzida pelo Google Tradutor;
- Tradução produzida pelo *DeepL*;
- Proposta de Tradução para português europeu;
- Em que motor se encontra o(os) erro(s) (Google Tradutor ou *DeepL*);
- Tipo de Erro MQM;
- Gravidade;
- Comentário.

De seguida, será exposto um exemplo retirado do corpus acima referido, para explicitação dos campos agora descritos:

Frase Original: “Great idea! I'll bring that up with him.”
 Frase Google Tradutor: “Boa ideia! Eu trago isso com ele.” – (1)
 Frase *DeepL*: “Ótima idéia! Vou falar sobre isso com ele.” – (2)

Motor de Tradução Automática	Tipo de Erro	Gravidade	Comentário
1. Google 2. <i>DeepL</i>	1. Accuracy - Mistranslation - Overly Literal 2. Fluency - Spelling - Diacritics	1. Critical 2. Minor	2. "ideia" não leva acento

Tabela 3- Exemplo da base de dados

Neste exemplo existem dois erros distintos. O primeiro erro (1) foi criado pelo Google Tradutor e foi classificado como crítico, pertencendo à categoria *Accuracy - Overly Literal* (Tradução Literal). O segundo erro (2) pertence à categoria *Diacritics* (Diacrítico), em *Fluency*, e foi catalogado como menor.

Após a anotação de todos os erros, foram criados campos extra de análise onde foi registado o número de erros por diálogo. Na análise realizada, foram primeiramente inseridos o número total de palavras em cada diálogo, assim como o número de erros distribuídos pelos dois motores. Estes erros foram categorizados como menor, grave ou crítico. Na tabela seguinte está representado um excerto da análise realizada:

Module ID	Diálogo	Nº de Turnos	Palavras Google Tradutor	Erro Menor	Erro Grave	Erro Crítico
Módulo 1	Diálogo 1	10	97	0	2	1
	Diálogo 2	10	105	0	2	0
	Diálogo 3	10	103	1	3	0
	Diálogo 4	10	94	1	1	1
	Diálogo 5	10	108	0	1	1

Tabela 4 – Exemplo dos campos extra de análise

4.3. Critérios de anotação de erros

Nesta secção, serão apresentados os critérios de anotação dos erros produzidos pelos motores de tradução automática, para português europeu, dos diversos diálogos, assim como a sua relevância para este estudo.

Para a análise e classificação dos erros foram utilizadas as métricas da MQM, apresentadas na secção 3.3. De maneira a ser feita uma análise mais detalhada dos tipos de erros, não foram empregues apenas as categorias principais/gerais da MQM Core, mas também as diversas subcategorias. Foram considerados apenas os erros das categorias *Accuracy* e *Fluency*, uma vez que é nestas que recai maioritariamente a inteligibilidade ou não das traduções.

Na categoria *Accuracy*, encontram-se todos os erros em que o texto de chegada não corresponde ao texto de partida. As subcategorias da categoria *Accuracy* são as seguintes:

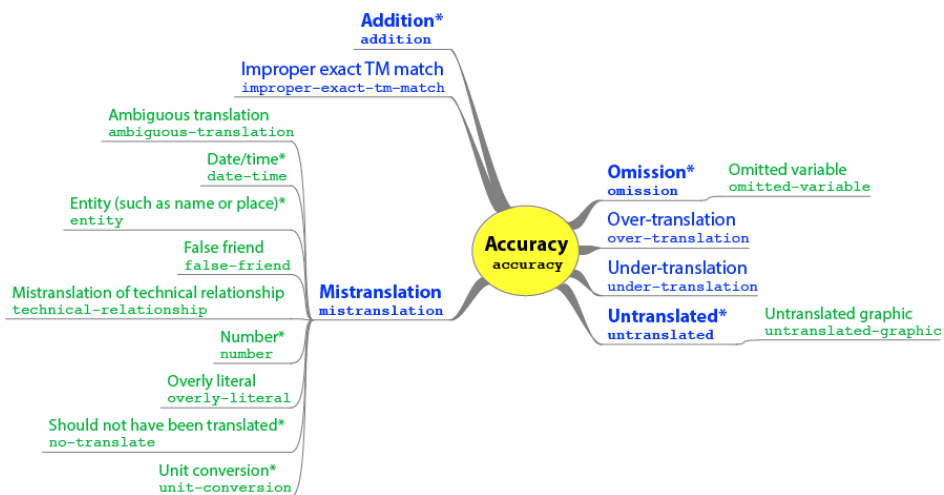


Figura 17 - Categoria Accuracy da MQM

Apesar de todas as subcategorias de *Accuracy* terem sido incluídas e analisadas neste estudo, a que será mais relevante é a *Mistranslation*, mais precisamente erros *Overly Literal*. Geralmente, erros desta tipologia incluem traduções literais, quase palavra a palavra, de frases, o que vai causar estranheza ao falante nativo por não ter a estrutura linguística correta. Outro exemplo desta tipologia de erro é a tradução incorreta de uma palavra, consoante o contexto em que esta está a ser utilizada. É expectável que seja nesta subcategoria que estejam incluídos maioritariamente os erros.

Na categoria *Fluency* encontram-se os erros que não permitem que o texto de chegada seja bem compreendido pelo falante, ou seja, todos os erros gramaticais. As suas subcategorias estão divididas da seguinte maneira:

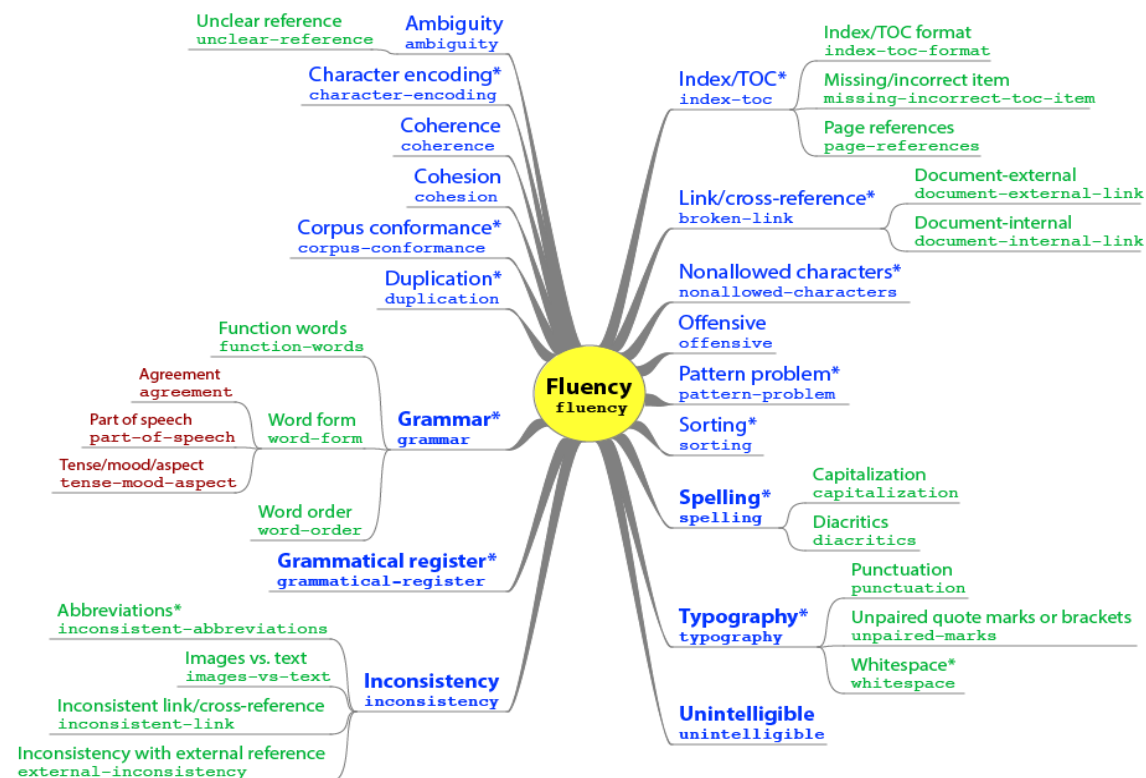


Figura 18 - Categoria Fluency da MQM

As tipologias de erros desta categoria consideradas mais relevantes para este estudo serão *Grammar* e *Coherence*. Para que qualquer texto seja fluente, é essencial que a gramática do mesmo esteja correta. Na subcategoria *Grammar*, vão estar todos os erros de concordância gramatical, erros relacionados aos modos e tempos verbais, à ordem das palavras, entre outros. Esta tipologia de erros vai influenciar grandemente a fluência do texto, pois afeta diretamente a sua compreensão. Outro erro significativo que danifica a fluência do texto é este não ser coerente no seu desenvolvimento. Erros deste tipo estão incluídos na subcategoria *Coherence*. No entanto, neste estudo serão também analisadas outras tipologias de erros que contribuem para a má compreensão do texto de chegada, produzido pelos motores de tradução automática Google Tradutor e *DeepL*.

A todos os erros é atribuído um nível de severidade, também utilizando as métricas MQM. O nível de severidade atribuído aos erros do Google Tradutor e do *DeepL* foi combinado

previamente com a empresa, em função do seu contexto e da sua importância. Foi realizada, por 3 anotadoras, (a mestrandas, a orientadora e a coorientadora deste relatório) a análise de uma amostra de 30 exemplos, que serviu de base para a discussão dos critérios sobre tipos de erros e graus de severidade.

Foi estabelecido que um erro crítico é um erro que muda completamente a semântica da frase ou um erro que resulta numa frase ininteligível para o falante. Por exemplo, a frase “I'd love to stay all summer, but I don't think I can swing it.” ser traduzida para “Adoraria ficar o verão todo, mas acho que não consigo balançar.” Neste exemplo, a tradução da frase, feita pelo *DeepL*, está incorreta, pois altera completamente a ideia principal da frase. A palavra “swing” foi traduzida literalmente, sem ter em consideração o contexto da frase. Este exemplo foi anotado como um erro de tradução literal. Uma tradução correta seria “Adorava ficar o verão todo, mas acho que não consigo.”

Um erro grave é um erro em que não há clareza na frase, o que poderá causar alguma estranheza a um falante nativo. Por exemplo, a frase “And how do you like these paintings?” foi traduzida pelo Google Tradutor para “E como você gosta dessas pinturas?”. Neste exemplo, a intenção da frase original é perguntar qual é a opinião de alguém em relação a uns quadros. Na frase traduzida, esta intenção ainda está presente, no entanto, por ser uma tradução literal do texto original, tem uma estrutura que parece atípica para um falante nativo. Foi considerado um erro de tradução literal e uma proposta de tradução seria “E o que achas destes quadros?”

Erros menores são todos os erros que envolvem os acentos ortográficos, como por exemplo, uma falta de acento ou um acento mal colocado, ou quando existe um espaço em branco a mais. Por exemplo, a frase “Great idea! I'll bring that up with him.” foi traduzida pelo *DeepL* para “Ótima idéia! Vou falar sobre isso com ele.”. Neste exemplo, existe um erro de acentuação, visto que a palavra ideia não leva acento agudo desde a entrada em vigor do novo acordo ortográfico, de janeiro de 2009. Este foi considerado um erro envolvendo um diacrítico.

É importante a atribuição de um nível de severidade, uma vez que isto vai permitir à empresa identificar onde existem problemas de tradução e, assim, poder resolvê-los de imediato. Localizar os erros críticos é fundamental, pois este tipo de erros leva a uma má compreensão do diálogo e prejudica a aprendizagem dos utilizadores da ELSA. É também

possível perceber quais os erros mais recorrentes e, no futuro, a empresa vai poder evitar a má tradução desses mesmos erros.

Dos tipos de erros das categorias *Accuracy* e *Fluency*, que serão explorados em detalhe no capítulo 5, vão estar englobados os erros mais recorrentes dos motores de Tradução Automática Google Tradutor e *DeepL*. É importante perceber que tipos de erros estes motores cometem com o objetivo de, no futuro, se treinar os sistemas de tradução automática a evitar estes erros e, por conseguinte, produzir traduções de melhor qualidade.

5. Resultados e Discussão

Neste capítulo, serão analisados e discutidos os erros anotados, apresentados brevemente na secção 4.3. Na secção 5.1, serão abordados os tipos de erros mais comuns dos motores de Tradução Automática Google Tradutor e *DeepL*, integrando na secção 5.1.1 os erros da categoria *Accuracy* e na secção 5.1.2 os erros relativos a *Fluency*.

5.1. Erros mais frequentes

Para uma adequada análise dos textos das traduções automáticas, é necessário compreender como se distribuem os erros anotados. Nos 186 diálogos analisados foram encontrados 448 erros no total, repartidos entre os dois motores. Será primeiramente evidenciado como estão estes erros distribuídos entre os dois motores. No gráfico seguinte, está representado o número total de erros catalogados do Google Tradutor e do *DeepL*:

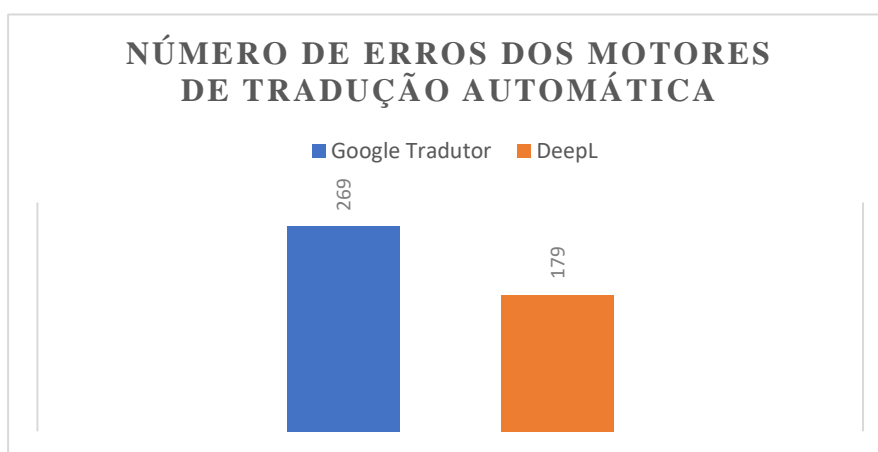


Gráfico 1 – Número de erros dos motores de tradução automática

Dos 448 erros anotados, 269 foram produzidos pelo Google Tradutor (60% dos erros) e 179 pelo *DeepL* (40%). Assim, é possível verificar que existe um número muito mais elevado de erros gerados pelo Google Tradutor do que pelo *DeepL*. Numa análise inicial, estes dados indicam que o motor de Tradução Automática *DeepL* produz menos erros e, assim, tem traduções de melhor qualidade.

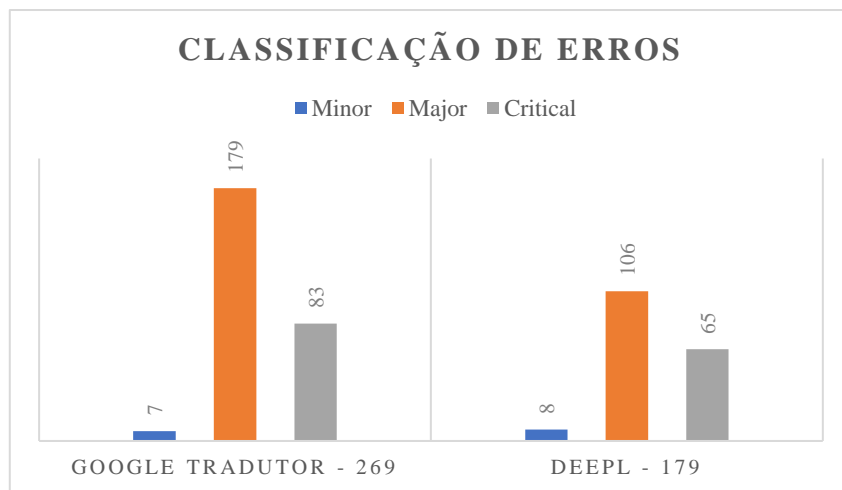


Gráfico 2 – Classificação de erros

Do número total de erros anotados, a maioria corresponde a erros de severidade grave ou crítica. 30% dos erros do Google Tradutor foram classificados como críticos, 67% como erros graves e apenas 3% são erros menores. Relativamente ao *DeepL*, 36% dos erros são críticos, 59% são erros graves e 5% dos erros são menores. É necessário referir que o número de erros críticos é ainda demasiado elevado.

Todos os erros foram agrupados nas duas categorias principais *Accuracy* e *Fluency*. O número total de erros nestas duas categorias, gerados pelos dois motores de tradução automática, está representado no seguinte gráfico:

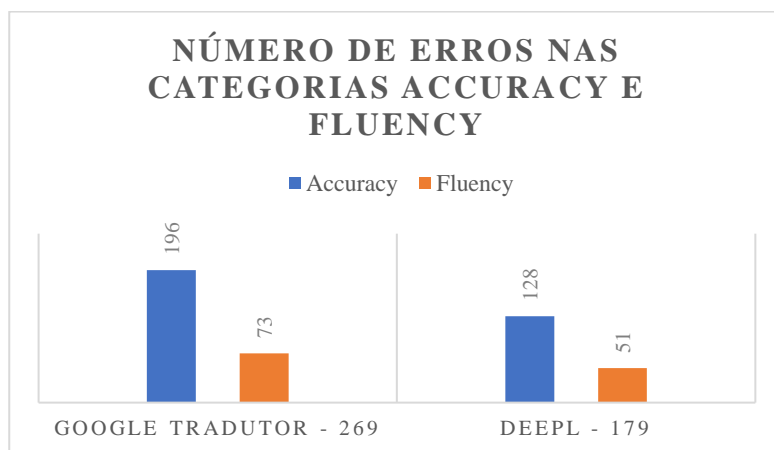


Gráfico 3 – Número de erros nas categorias Accuracy e Fluency

Tanto no motor de tradução automática Google Tradutor, como no *DeepL* foram encontrados mais erros na categoria *Accuracy*. Dos 269 erros do motor de tradução automática Google Tradutor, 196 estão catalogados na categoria *Accuracy* (73%) e 73 em

Fluency (27%). Já em relação ao *DeepL*, é possível verificar que, dos 179 erros, 128 estão atribuídos a *Accuracy* (72%) e 51 a *Fluency* (28%).

5.1.1. Distribuição de erros na categoria *Accuracy*

Na categoria *Accuracy*, foram anotados um número total de 324 erros, sendo esta a categoria em que recaem maioritariamente os erros. Nas frases do Google Tradutor foram encontrados 196 erros (60% dos erros) e nas do *DeepL* 128 (40%). Estes erros foram organizados em diferentes subcategorias da MQM que estão graficamente distribuídas da seguinte maneira:

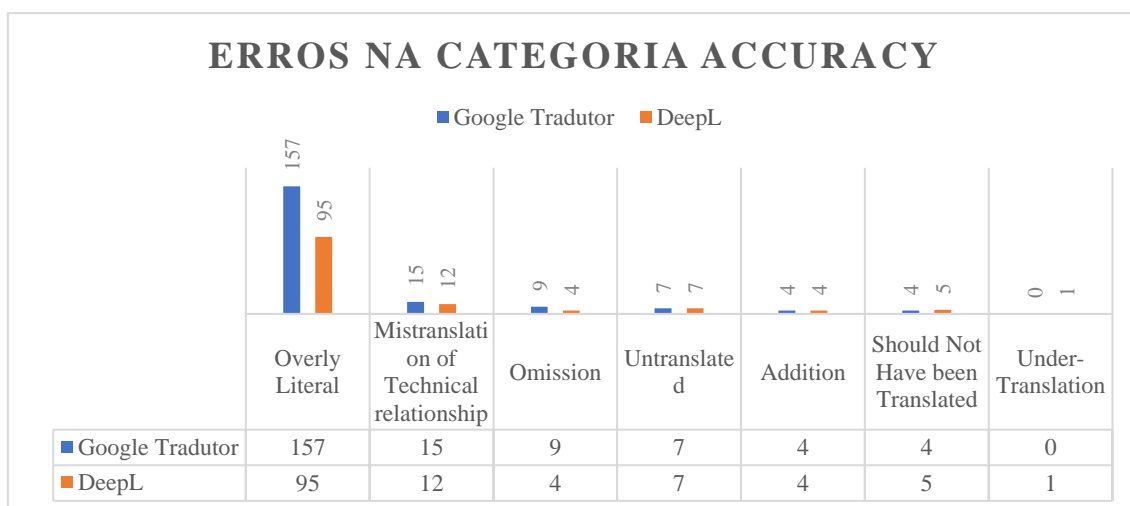


Gráfico 4 – Erros na categoria *Accuracy*

Nas subsecções seguintes, serão analisadas as subcategorias MQM em que foram encontrados estes erros e serão também expostos alguns exemplos de erros e um breve comentário sobre os mesmos.

5.1.1.1. *Overly Literal* (Tradução Literal)

Nesta subcategoria estão todos os erros que sejam fruto de uma tradução literal do texto original, criando assim uma frase estranha ao falante nativo, seja por ter sido traduzido palavra a palavra, ou por ter uma palavra que não está a ser utilizada no contexto correto. Foram anotados 252 erros, dos quais 157 pertencem ao Google Tradutor (62% dos erros)

e 95 ao *DeepL* (38%), sendo este o tipo de erro mais encontrado neste estudo. O maior obstáculo deste tipo de tradução, para o falante nativo, é o de as frases não terem a estrutura linguística expectável num determinado contexto e/ou adequada e, por isso, soarem invulgares. Na estrutura MQM, erros *Overly Literal* estão incluídos na subcategoria *Mistranslation*. Exemplos desta tipologia de erros são:

1. (i) Good point. I'll start looking for some hotels.
- (ii) Bom ponto. Vou começar a procurar alguns hotéis. (Frase Google Tradutor)
- (iii) Ponto bom. Vou começar a procurar alguns hotéis. (Frase *DeepL*)
- (iv) Bem visto. Vou começar a procurar alguns hotéis. (Tradução proposta)

No exemplo 1, tanto o motor de Tradução Automática Google Tradutor como o *DeepL* traduziram “Good Point” literalmente. Em português europeu, esta tradução não é natural ao falante nativo. Apesar de o significado desta expressão ser perceptível, não é algo que fosse produzido pelo falante nativo naquele contexto. Assim, este erro foi classificado como grave.

2. (i) Of course! Let's talk about something even funnier tomorrow: tongue twisters!
- (ii) Claro que sim! Vamos falar de algo ainda mais engraçado amanhã: tornados-língua! (Frase *DeepL*)
- (iii) Claro! Amanhã vamos falar de algo ainda mais engraçado: trava-línguas. (Tradução proposta)

No exemplo 2, existe uma clara tradução literal, do *DeepL*. Nesta frase, o motor de tradução automática não considerou “tongue twisters” como uma única expressão, mas sim duas palavras completamente independentes. “Tornados-língua” não existe no léxico português e, por isso, esta tradução literal vai impedir a compreensão plena da frase. Assim, o erro foi considerado crítico.

3. (i) Sounds like a good catch.
- (ii) Parece uma boa captura. (Frase Google Tradutor)
- (iii) Parece um bom partido. (Tradução proposta)

No exemplo 3, apresenta-se a expressão “good catch”. Nesta frase, o Google Tradutor não consegue identificar a expressão e, por esse motivo, recorre à tradução literal. Na

língua portuguesa, “boa captura” não é adequado para este contexto. Assim, este é um erro crítico.

É na subcategoria *Overly Literal* que estão compreendidos o maior número de erros, tanto do motor de tradução automática Google Tradutor, como do *DeepL*. Assim, pode afirmar-se que ambos os motores não têm ainda treino de contexto suficiente que lhes permita compreender o contexto das frases apresentadas e, por esse motivo, traduzem certas palavras/expressões inadequadamente.

5.1.1.2. *Mistranslation of Technical Relationship* (Erro de tradução de natureza técnica)

Nesta subcategoria estão todos os erros que contenham um conjunto de palavras conectadas ou expressões que, apesar de parecerem plausíveis, não são consideradas corretas na língua de chegada, devido ao conhecimento linguístico inerente ao falante nativo. Foram encontrados 27 erros desta tipologia. 15 desses erros foram gerados pelo Google Tradutor (56%) e 12 pelo *DeepL* (44%). Estes erros encontram-se na subcategoria *Mistranslation* na tipologia da MQM. Alguns exemplos englobados nesta subcategoria são os seguintes:

4. (i) Thank you for grabbing this table.
- (ii) Obrigado por pegar esta mesa. (Frase Google Tradutor)
- (iii) Obrigado por teres agarrado esta mesa. (Frase *DeepL*)
- (iv) Obrigado por teres arranjado mesa. (Tradução proposta)

No exemplo 4, apesar de ambas as frases traduzidas serem gramaticalmente corretas, não existe uma relação correta neste contexto. No dia-a-dia, no contexto de querer ter uma mesa, não se utiliza nem o verbo “pegar” nem o verbo “agarrar”. Embora não seja considerada uma expressão adequada, é possível entender o significado da frase que é “ter uma mesa”. Por este motivo, foi considerado um erro grave.

5. (i) You're sweet. Let's go have some hot chocolate together.
- (ii) És um querido. Vamos comer chocolate quente juntos. (Frase *DeepL*)

(iii) És uma querida. Vamos beber chocolate quente juntas. (Tradução proposta)

No exemplo 5, o *DeepL*, associou “comer” com “chocolate quente”. Chocolate quente é considerado uma bebida e, por isso, não é algo que se possa comer, mas sim beber. Assim, não deve existir uma relação entre as duas palavras. Este erro foi catalogado como grave, visto que ainda é possível inferir o significado da frase.

6. (i) Let’s drink a toast! Happy New Year, everybody!

(ii) Vamos beber um brinde! Feliz Ano Novo a todos! (Frase Google Tradutor)

(iii) Vamos fazer um brinde! Feliz Ano Novo a todos! (Tradução proposta)

No exemplo 6, o motor de tradução automática Google Tradutor relacionou “beber” a “brinde”. No português europeu, não está correto dizer que se “bebe um brinde”, mas sim, que se “faz um brinde”. Uma vez que o sentido da frase é ainda compreensível, este erro foi considerado grave.

5.1.1.3. *Omission* (Omissão):

Quando partes da informação do texto de partida não está presente no texto de chegada, o que leva a que o leitor não tenha acesso ao texto original completo. A esta subcategoria foram adicionados 9 erros do Google Tradutor (69%) e 4 erros do *DeepL* (31%), tendo na totalidade 13 erros. De seguida serão apresentados alguns exemplos:

7. (i) Yes, I will text you them when I get home.

(ii) Sim, enviarei uma mensagem quando chegar em casa. (Frase Google Tradutor)

(iii) Sim, vou enviar-lhe um SMS quando chegar a casa. (Frase *DeepL*)

(iv) Sim, eu envio-tas por mensagem quando chegar a casa. (Tradução proposta)

No exemplo 7, tanto o Google Tradutor como o *DeepL* ocultaram uma informação importante na sua tradução: que ia ser enviado algo para a outra pessoa na mensagem. A informação presente nas duas traduções é a de que alguém vai enviar uma mensagem a outra pessoa quando chegar a casa. No entanto, na frase original, está expresso que alguém vai enviar algo por mensagem, quando chegar a casa. Assim, as traduções do

Google Tradutor e do *DeepL* não correspondem totalmente à frase original. Este exemplo foi considerado um erro grave.

8. (i) Hey, was that your brother I met at the party yesterday?
(ii) Ei, foi seu irmão que conheci ontem? (Frase Google Tradutor)
(iii) Olá, foi o teu irmão que eu conheci ontem na festa? (Tradução proposta)

No exemplo 8, a tradução do Google Tradutor está a ocultar a informação de onde os sujeitos da frase se conheceram. No entanto, a informação principal da frase (se os sujeitos se conheceram ontem) está evidente na tradução e, por isso, este foi considerado um erro grave, apesar de a frase estar incompleta.

9. (i) Maybe I can show you around if your plans line up with mine. How long would you be staying?
(ii) Talvez eu possa te mostrar se seus planos estão alinhados com os meus. Quanto tempo você ficaria? (Frase Google Tradutor)
(iii) Talvez eu possa te mostrar se seus planos se alinham com os meus. Quanto tempo você ficaria? (Frase *DeepL*)
(iv) Talvez te possa fazer uma visita guiada se tivermos os mesmos horários. Quanto tempo ficarias? (Tradução proposta)

No exemplo 9, ambos os motores de tradução automática omitiram o que iria ser mostrado pelo sujeito e, por conseguinte, não está a ser transportada pela tradução toda a informação necessária para a compreensão total da frase. Este é um erro crítico.

Na subcategoria *Omission* não foram encontrados muitos erros o que indica que esta tipologia de erros não é muito comum nas traduções do Google Tradutor e do *DeepL*.

5.1.1.4. *Untranslated* (Não traduzido):

Nesta subcategoria foram apontados todos os erros em que uma frase/expressão/palavra não tenha sido traduzida, fazendo com que o leitor não compreenda totalmente o texto

traduzido. Estão inseridos 14 erros nesta categoria, sendo 7 do Google Tradutor (50%) e 7 do *DeepL* (50%). De seguida serão apresentados alguns exemplos:

10. (i) That's a no brainer! My friends and I are going for sure.
(ii) Isso é um "no brainer"! Meus amigos e eu vamos com certeza. (Frase *DeepL*)
(iii) Óbvio! Eu e os meus amigos vamos ver de certeza. (Tradução proposta)

No exemplo 10, o motor de tradução automática *DeepL* não traduziu a expressão “no brainer”, o que leva a o leitor não seja capaz de entender totalmente o que está a ser apresentado na frase. Este erro foi classificado como crítico, uma vez que vai afetar a compreensão do leitor.

11. (i) The hustling atmosphere and the excellent amenities
(ii) A atmosfera de hustling e as excelentes amenidades. (Frase *DeepL*)
(iii) O ambiente agitado e as excelentes comodidades. (Tradução proposta)

No exemplo 11, não foi traduzida a palavra “hustling”. Deste modo, a frase não faz qualquer sentido e o leitor não consegue compreender devidamente o significado da frase. Por este motivo, este erro foi classificado como crítico.

12. (i) I will make the stuffing and cranberry sauce as well. It will be delicious!
(ii) Vou fazer o recheio e molho de cranberry também. Vai ser delicioso! (Frase Google Tradutor)
(iii) Também vou fazer o recheio e o molho de arando. Vai ficar delicioso! (Tradução proposta)

No exemplo 12, não foi traduzido o ingrediente principal do molho. Este erro foi considerado grave, uma vez que a ideia principal da frase (que o sujeito fará o recheio e o molho) está devidamente traduzida.

Não foi encontrado um número significativo de erros nesta subcategoria, o que sugere que erros desta tipologia não são frequentes nos dados analisados.

5.1.1.5. *Addition* (Adição)

Nesta categoria estão todos os erros em que exista, no texto de chegada, informação adicional que não está presente no texto de partida. Nesta subcategoria foram incluídos 8 erros, dos quais 4 (50%) provocados pelo Google Tradutor e 4 (50%) pelo *DeepL*.

Exemplos desta tipologia de erros são:

13. (i) Luckily, I don't get sick when it goes from hot to cold! Haha
(ii) Felizmente, não fico doente quando fica quente ou frio! (risadinha) (Frase Google Tradutor)
(iii) Felizmente, eu não fico doente quando passa de quente a frio! Haha (Tradução proposta)

No exemplo 13, o Google Tradutor traduziu incorretamente parte da frase, introduzindo informação que não existe na frase original. Na frase em inglês, a mensagem é a de que o sujeito não fica doente quando passa de uma temperatura quente para uma temperatura mais fria. No entanto, na frase traduzida a informação apresentada é a de que o sujeito nunca fica doente quando está quente ou frio. Esta adição de informação inexata leva a que os leitores tenham uma compreensão incorreta da frase. Este erro foi considerado crítico, uma vez que o que está a ser apresentado na frase traduzida não corresponde à frase original.

14. (i) Pretty good, I charge my phone every other day.
(ii) Muito bom, eu carrego meu telefone todos os dias. (Frase Google Tradutor)
(iii) Muito bem, eu carrego o meu telefone de dois em dois dias. (Frase *DeepL*)
(iv) Bastante boa, eu carrego o meu telemóvel dia sim dia não. (Tradução proposta)

No exemplo 14, a expressão “every other day” foi erradamente traduzida para “todos os dias” pelo Google Tradutor e “de dois em dois dias” pelo *DeepL*. Na frase original, o sujeito refere que carrega o seu telemóvel dia sim dia não, no entanto, nas frases traduzidas foi adicionada informação incorreta. Neste exemplo, foi acrescentada

informação que não existe na frase original e, por este motivo, este erro foi considerado crítico.

15. (i) I drink boba five times a week! By the way, what are your weekend plans?

(ii) Eu bebo boba cinco vezes por semana! Pela maneira, que são suas plantas do fim de semana? (Frase do *DeepL*)

(iii) Eu bebo boba cinco vezes por semana! Por falar nisso, quais são os teus planos para o fim de semana? (Tradução proposta)

No exemplo 15, surge na tradução do *DeepL* a palavra “planta”. Na frase original não existe qualquer referência a planta, mas sim “plans”, que traduzida significa “planos”. Neste exemplo, o motor de tradução automática traduziu uma palavra completamente errada e no contexto incorreto. Assim, este erro foi classificado como crítico.

Nesta subcategoria não estão incluídos um avultado número de erros. Assim, pode afirmar-se que erros de *Addition* não são normalmente encontrados nas traduções dos motores automáticos Google Tradutor e *DeepL*.

5.1.1.6. *Should Not Have been Translated* (Não deveria ter sido traduzido)

Foram considerados todos os erros onde, no texto de chegada, uma determinada palavra/expressão não deveria ter sido traduzida, pois leva à má compreensão da informação apresentada. Nesta subcategoria estão inseridos 9 erros, sendo 4 destes erros criados pelo Google Tradutor (44% dos erros) e 5 pelo *DeepL* (56%). Alguns exemplos são:

16. (i) She's in the Glow is my favorite one.

(ii) Ela está no brilho é a minha favorita. (Frase Google Tradutor)

(iii) Ela está no Glow é a minha favorita. (Frase *DeepL*)

(iv) "She's in the Glow" é o meu favorito. (Tradução proposta)

No exemplo 16, “She’s in the Glow” é o nome de um *blog*. Visto que é considerado nome próprio, não deveria ser traduzido, uma vez que se for, o leitor não irá encontrar qualquer

informação sobre o mesmo, caso o procure *online*. Por este motivo, este erro foi considerado crítico.

17. (i) Dreaming in Blush is really great too.
(ii) Sonhar em Blush também é ótimo. (Frase Google Tradutor)
(iii) Sonhar em Blush também é muito bom. (Frase *DeepL*)
(iv) "Dreaming in blush" também é muito bom. (Tradução proposta)

No exemplo 17, tanto o Google Tradutor como o *DeepL*, traduziram para português “Dreaming in Blush”, o nome de um *blog*. Este erro foi considerado crítico.

18. (i) I’m using Windows 8 on my laptop.
(ii) Estou a usar a Janela 8 do meu portátil. (Frase *DeepL*)
(iii) Estou a usar o Windows 8 no meu computador portátil. (Tradução proposta)

No exemplo 18, foi traduzido o nome do sistema operativo “Windows 8”. Este sistema é conhecido em todo o mundo pelo seu nome original e por isso não deve ser traduzido, pois existe o risco de o leitor não compreender o que está a ser referido. Assim, este erro foi classificado como crítico.

Estes erros indicam que os motores de tradução automática têm ainda alguma dificuldade em distinguir nomes próprios de nomes comuns e, por esse motivo, os traduzem.

5.1.1.7. *Under-Translation* (Traduzido de forma incompleta):

É considerado erro quando o texto de chegada não contém informação tão específica como no texto de partida. Nesta subcategoria está incluído 1 erro apenas, gerado pelo motor de tradução automática *DeepL*.

19. (i) What is Star Wars about?
(ii) O que é Star Wars? (Frase *DeepL*)
(iii) Star Wars é sobre o quê? (Tradução proposta)

No exemplo 19, a frase original questiona sobre o que é Star Wars, subentendendo que o sujeito entende que é um filme, mas não conhece o seu enredo. No entanto, na tradução

do *DeepL*, o que está descrito é que o sujeito não conhece de todo o que é *Star Wars*, ou seja, não sabe que é um filme, nem qual é o seu enredo. Neste exemplo, a tradução é bastante mais generalizada do que a frase original. Assim, este é um erro grave.

Posto que apenas foi considerado 1 erro nesta subcategoria, pode afirmar-se que esta tipologia de erros é muito rara nos dados analisados.

A generalidade dos erros encontrados neste estudo estão inseridos na categoria principal *Accuracy*. Assim, é possível depreender que a maior parte dos erros anotados está diretamente relacionada à informação incorretamente traduzida pelos motores de tradução automática Google Tradutor e *DeepL*, o que vai induzir dificuldades na compreensão do texto.

5.1.2. Distribuição de erros na categoria *Fluency*

A segunda categoria principal da MQM, que será nesta subsecção desenvolvida, é *Fluency*. Nesta categoria estão incluídos todos os erros que não vão permitir a compreensão do texto de chegada, devido a variados erros gramaticais. Foram englobados 124 erros, 73 criados pelo Google Tradutor (59% dos erros) e 51 pelo *DeepL* (41%). Estes erros estão representados no seguinte gráfico:

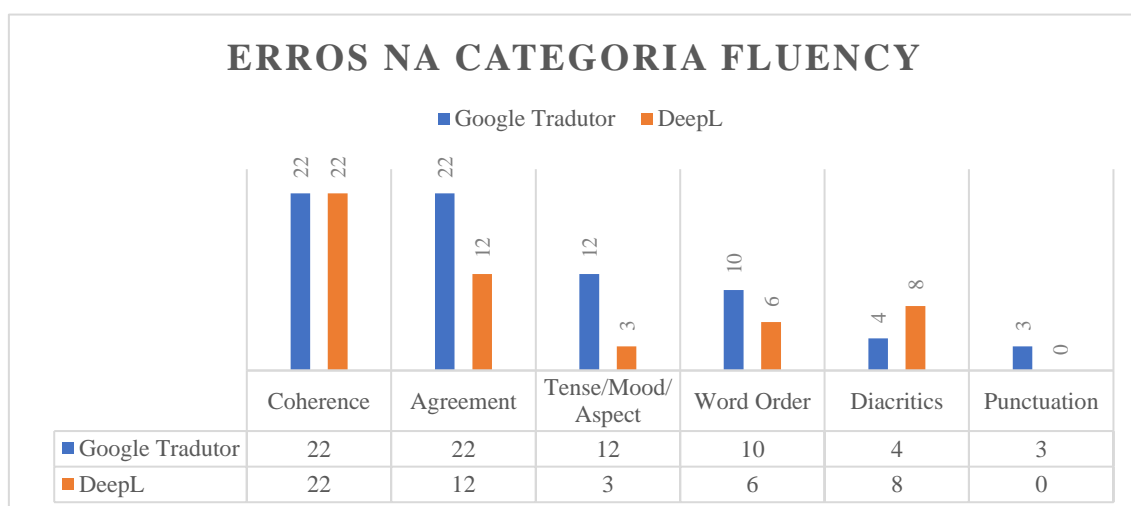


Gráfico 5 – Erros na categoria *Fluency*

Nas subseções seguintes serão analisadas as subcategorias de *Fluency*, complementadas por alguns exemplos de erros anotados e um breve esclarecimento dos mesmos.

5.1.2.1. *Coherence* (Coerência)

Nesta subcategoria estão os erros que não permitem que o texto seja coerente. Foram inseridos 44 erros de *Coherence*. 22 destes erros foram produzidos pelo Google Tradutor (50%) e 22 pelo *DeepL* (50%). Alguns exemplos de erros encontrados nesta subcategoria são:

20. (i) Watching football is a social activity. You watch it with friends and family.
(ii) Ver futebol é uma atividade social. Você assiste com amigos e familiares. (Frase *DeepL*)
(iii) Ver futebol é uma atividade social. Vê-se com amigos e família. (Tradução proposta)

No exemplo 20, o *DeepL* traduz primeiramente a palavra “Watching” para “Ver” e, de seguida, traduz a mesma palavra para “assistir”. Aqui existe um problema de coerência, uma vez que a mesma palavra foi traduzida de duas maneiras diferentes num mesmo diálogo. No entanto, o significado de ambas é equivalente e, por isso, este erro foi classificado como grave.

21. (i) How many users do you have?
(ii) Quantos usuários você tem? (Frase *DeepL*)
(iii) We have eight thousand registered users.
(iv) Temos oito mil *utilizadores* registados. (Frase *DeepL*)
(v) Temos oito mil utilizadores registados. (Tradução proposta)

No exemplo 21, a palavra “users” é traduzida para usuários, uma palavra maioritariamente utilizada por falantes do português do Brasil e, numa segunda frase do mesmo diálogo, “users” é traduzido pelo mesmo motor de tradução automática para “utilizadores”, palavra do português europeu. Uma vez que o significado de ambas as

traduções é o mesmo e o sentido da frase ainda é evidente, foi decidido que este erro é grave.

22. (i) Piano sounds fun. I've been practicing a lot of guitar recently.
(ii) Piano parece divertido. Tenho praticado muito violão recentemente. (Frase *DeepL*)
(iii) That's cool. How long have you been playing?
(iv) Isso é fixe. Há quanto tempo estás a jogar? (Frase *DeepL*)
(v) Isso é fixe. Há quanto tempo tocas? (Tradução proposta)

No exemplo 22, a temática do diálogo são instrumentos musicais. Na frase (i), o sujeito refere que tem estado a praticar violão e o *DeepL* faz uma tradução adequada da frase, em (ii). No entanto, na frase (iii), a tradução da mesma é completamente descontextualizada, uma vez que o motor de tradução automática traduz a palavra “playing” para “jogar” em vez de “tocar”, que seria a tradução correta para este contexto. Este erro foi classificado como crítico, visto que a tradução do *DeepL* não tem o mesmo significado semântico que a frase original.

Foram encontrados alguns erros significativos desta tipologia. A maioria destes erros deve-se ao Google Tradutor e ao *DeepL* traduzirem automaticamente frases separadas, não tendo em conta que várias frases podem pertencer a um mesmo diálogo. Por esse motivo, pode ser primeiramente utilizada a tradução de uma palavra numa frase e, de seguida, essa mesma palavra é traduzida de maneira diferente. Outra razão para erros desta tipologia pode ser o de na língua inglesa, uma mesma palavra poder ter vários significados, consoante o contexto em que esta está a ser utilizada, ao contrário da língua portuguesa que é mais específica nas suas palavras. Assim, os motores de tradução automática, não tendo em conta várias frases para compreender o contexto, podem promover traduções incorretas.

5.1.2.2. *Agreement* (Concordância Gramatical)

Os erros desta subcategoria correspondem a uma ou mais palavras que não concordam em género ou número com palavras de frases anteriores, de um mesmo diálogo. Dentro da categoria principal, *Fluency*, *Agreement* está estruturalmente incluída na subcategoria

Grammar (Gramática) → *Word Form* (Forma das Palavras) → *Agreement*. Foram adicionados 34 erros deste tipo, sendo 22 do Google Tradutor (65% dos erros) e 12 do *DeepL* (35%). Alguns exemplos são:

23. (i) Would you like to bring your delicious pumpkin pies again? Everyone loved them last year.
- (ii) Gostaria de trazer suas deliciosas tortas de abóbora novamente? Todo mundo os amou no ano passado (Frase Google Tradutor)
- (iii) Queres trazer as tuas deliciosas tartes de abóbora outra vez? Toda a gente os adorou no ano passado. (Frase *DeepL*)
- (iv) Queres trazer as tuas deliciosas tartes de abóbora outra vez? Toda a gente as adorou no ano passado. (Tradução proposta)

No exemplo 23, tanto na tradução do Google Tradutor, como na tradução do *DeepL*, existe uma incorreta concordância gramatical com a palavra “torta”. Ambos os motores de tradução não identificam que o pronome da segunda frase se referia à palavra “torta” e que, por esse motivo, o pronome deveria estar no feminino. Uma vez que a ideia principal da frase ainda é compreensível, este erro foi classificado como grave.

24. (i) Yes, I love my sisters. They are kind and smart.
- (ii) Sim, eu amo minhas irmãs. Eles são gentis e inteligentes. (Frase Google Tradutor)
- (iii) Sim, eu amo as minhas irmãs. Eles são bondosos e inteligentes. (Frase *DeepL*)
- (iv) Sim, amo as minhas irmãs. São amáveis e inteligentes. (Tradução proposta)

No exemplo 24, ambos os motores de tradução automática utilizaram o pronome errado. A palavra “irmãs” é um nome feminino e, por isso, deveria ter sido utilizado o pronome “elas”, em vez de “eles”. Este erro foi classificado como grave.

25. (i) I see. My favorite animal is turtles. They are so cute.
- (ii) Eu vejo. Meu animal favorito é tartarugas. Eles são tão bonitinhos. (Frase Google Tradutor)
- (iii) Estou a ver. O meu animal favorito são as tartarugas. São tão giros. (Frase *DeepL*)
- (iv) Estou a ver. O meu animal favorito é a tartaruga. Elas são tão queridas. (Tradução proposta)

No exemplo 25, o pronome que deveria ter sido utilizado tanto pelo Google Tradutor, como pelo *DeepL*, é “elas” uma vez que “tartarugas” é um nome feminino. Para além do uso incorreto do pronome, existe também um erro na concordância gramatical entre o nome e o adjetivo, que deveria estar no feminino. Este foi considerado um erro grave, uma vez que o conteúdo da frase é ainda compreensível.

5.1.2.3. *Tense/Mood/Aspect* (Tempos e Modos Verbais)

Nesta subcategoria estão todos os erros que contenham um verbo, modo ou tempo verbal incorreto. Na estrutura MQM, esta subcategoria encontra-se em *Grammar* (Gramática) → *Word Form* (Forma das palavras) → *Tense/Mood/Aspect*. Foram encontrados 15 erros deste tipo, 12 pertencentes ao Google Tradutor (80% dos erros anotados) e 3 ao *DeepL* (20%). De seguida, serão apresentados alguns exemplos de erros:

26. (i) It's so nicely wrapped.
(ii) É tão bem embrulhado. (Frase Google Tradutor)
(iii) Está tão bem embrulhado. (Tradução proposta)

No exemplo 26, o Google Tradutor traduziu incorretamente o verbo “to be”. Em português, este verbo pode ser traduzido de duas maneiras distintas, “ser” ou “estar”, dependendo do contexto em que está a ser utilizado. Neste exemplo, não poderia ser utilizado o verbo “ser”, uma vez que não vai formular uma frase gramaticalmente correta. No entanto, este erro foi classificado como grave, pois a noção de que “algo está bem embrulhado” é ainda perceptível ao leitor.

27. (i) It was lovely my granddad and I went ice fishing.
(ii) Foi adorável meu avô e eu fui pescar no gelo. (Frase Google Tradutor)
(iii) Foi maravilhoso, eu e o meu avô fomos pescar no gelo. (Tradução proposta)

No exemplo 27, o verbo não foi conjugado corretamente. Existem dois sujeitos na frase e, por isso, o verbo “ir” tem obrigatoriamente de estar conjugado na 1ª pessoa do plural. Este erro foi classificado como grave.

Erros desta tipologia não são muito comuns, no entanto, parece haver maior incidência nas traduções do Google Tradutor.

5.1.2.4. *Word Order* (Ordem das palavras)

Nesta subcategoria encontram-se os erros onde a ordem das palavras não está correta e, por esse motivo, o texto vai soar estranho a um falante nativo. Nesta subcategoria estão inseridos 16 erros, sendo 10 destes erros do motor de tradução automática Google Tradutor (63%) e 6 do *DeepL* (37%). Alguns exemplos são:

28. (i) Haha. Perfect. I'll let everyone know about the next one soon.
(ii) Haha Perfeito. Vou deixar todos saberem sobre o próximo em breve. (Frase Google Tradutor)
(iii) Haha. Perfeito. Brevemente irei marcar com todos o próximo jantar. (Tradução Proposta)

No exemplo 28, a disposição da frase produzida pelo motor de tradução automática Google Tradutor vai causar estranheza ao falante nativo de português europeu. Este erro foi catalogado como grave.

29. (i) Are visits from the CEO usually announced?
(ii) As visitas do CEO geralmente são anunciadas? (Frase Google Tradutor)
(iii) As visitas do CEO são normalmente anunciadas? (Tradução Proposta)

No exemplo 29, o advérbio “geralmente” está colocado, pelo Google Tradutor, imediatamente antes do verbo, o que não é uma prática comum na língua portuguesa. Este advérbio deveria ser utilizado ou no início da frase ou no fim.

30. (i) How do I politely leave a dinner party?
(ii) Como educadamente saio de um jantar? (Frase Google Tradutor)

(iii) Como é que saio educadamente de um jantar? (Tradução proposta)

No exemplo 30, “educadamente” deveria estar presente depois do verbo “sair”. Na tradução do Google Tradutor, a pergunta parece estranha e incompleta a um falante nativo de português europeu. No entanto, como a pergunta ainda é compreensível, este erro foi considerado grave.

5.1.2.5. *Diacritics* (Diacríticos)

São erros relativos a acentos gráficos mal colocados. Esta subcategoria está inserida em *Spelling* (Ortografia) → *Diacritics* e nela estão incluídos 12 erros. O Google Tradutor gerou 4 destes erros (33%) e o *DeepL* 8 erros 67%). Exemplos desta tipologia são:

31. (i) I'm still looking for ideas. Any recommendations?
(ii) Eu ainda estou procurando idéias. Alguma recomendação? (Frase Google Tradutor)
(iii) Ainda estou à procura de ideias. Alguma recomendação? (Tradução proposta)
32. (i) Great idea! I'll bring that up with him.
(ii) Ótima idéia! Vou falar sobre isso com ele. (Frase DeepL)
(iii) Boa ideia! Vou falar disso com ele. (Tradução proposta)
33. (i) You know, that's not a bad idea. I can put together a group and invite you.
(ii) Você sabe, isso não é uma má idéia. Posso reunir um grupo e convidá-lo. (Frase DeepL)
(iii) Sabes, isso não é uma má ideia. Eu posso criar um grupo e convidar-te. (Tradução proposta)

Todos os erros relacionados a acentos gráficos são referentes à palavra “ideia”. Em algumas frases, os motores de tradução automática parecem não reconhecer o acordo ortográfico de 1990 e, por isso, utilizam ainda o acento agudo. No entanto, este acento é ainda utilizado no Brasil o que indica que os motores de tradução automática podem estar a produzir traduções seguindo regras gramaticais do português do Brasil. Todos os erros

desta subcategoria foram classificados como menores, uma vez que não afetam de maneira nenhuma a compreensão do texto.

5.1.2.6. *Punctuation* (Pontuação)

Estão nesta subcategoria todos os erros onde ocorra um erro de pontuação, que envolvam por exemplo uma vírgula ou um ponto final. Existem 3 erros desta tipologia, todos eles correspondentes ao Google Tradutor. Alguns exemplos são:

34. (i) Ah, I see. I guess he's just trying to be fair.
(ii) Ah entendo. Eu acho que ele está apenas tentando ser justo. (Frase Google Tradutor)
(iii) Ah, estou a perceber. Acho que ele está apenas a tentar ser justo. (Tradução proposta)

No exemplo 34, não está assinalada a vírgula após a interjeição “Ah”. Este erro foi classificado como menor.

35. (i) People usually eat brunch between 10 A.M. and 2 P.M.
(ii) As pessoas costumam comer um brunch entre as 10 da manhã e 2 P.M. (Frase Google Tradutor)
(iii) Normalmente, as pessoas costumam comer brunch entre as 10h e as 14h. (Tradução proposta)

No exemplo 35, existe um ponto final incorretamente colocado após “manhã”. Foi considerado um erro menor.

Erros desta tipologia não são considerados cruciais para a compreensão do texto de chegada e, por isso, são normalmente considerados erros menores.

Os erros apresentados nas secções 5.1.1 e 5.1.2 foram considerados os mais relevantes nas traduções do Google Tradutor e do *DeepL*. Depois da análise de todos os erros, pode concluir-se que nenhum destes motores é ainda capaz de produzir traduções

completamente corretas, uma vez que existe algum tipo de erro, que foi avaliado como menor, grave ou crítico, em muitas das frases examinadas neste estudo.

6. Conclusão e Trabalho futuro

Neste estudo pode concluir-se que ambos os motores de Tradução Automática, Google Tradutor e *DeepL*, cometem ainda alguns erros de tradução e que, por este motivo, requerem alguma pós-edição humana. Dos 186 diálogos examinados, exercícios da aplicação *ELSA Speak*, foram registados 448 erros que foram organizados nas diferentes

categorias de erros da tipologia MQM – a mais utilizada atualmente para a classificação da qualidade das traduções.

Entre estes erros encontram-se maioritariamente traduções literais e erros de coerência textual. Contudo, foram também registados erros de omissões, adições, modos e tempos verbais, diacríticos, entre outros. Uma vez que nos exemplos analisados do motor de tradução automática *DeepL* foram encontrados menos erros do que nos do Google Tradutor, pode afirmar-se que este é capaz de produzir traduções mais corretas e exatas, não podendo, no entanto, ser ainda comparado a um tradutor humano. Ambos os motores de tradução produzem ainda inúmeros erros.

A realização deste estudo terá múltiplas contribuições para trabalhos futuros. Será particularmente importante para a empresa ELSA Corp. e a sua aplicação *ELSA Speak*, pois vai possibilitar a identificação de exercícios que poderão ter um maior número de erros de tradução e, assim, retificá-los de imediato, especialmente os erros críticos que levam à má compreensão dos exercícios. É de grande importância para a empresa evitar este tipo de erro de tradução na aplicação, pois pode levar a uma eventual advertência de um utilizador.

Uma vez que o Google Tradutor e o *DeepL* são utilizados diariamente por milhares de utilizadores, é também importante entender que tipo de erros estes motores de tradução automática cometem, para que, no futuro, estes sistemas NMT possam aprender a produzir traduções aperfeiçoadas e que exijam menos pós-edição humana.

Bibliografia

AA.VV., Discurso e Marcadores Discursivos. Em Gramática do português, vol.III, cap. 51. Lisboa: Fundação Calouste Gulbenkian, agosto de 2020.

ALPAC. 1966. *Language and Machines: Computers in Translation and Linguistics: A Report*. Washington: National Academy of Sciences, National Research Council.

AMTA. 2016. Proceedings of AMTA 2016, vol1: MT Researcher's Track. Austin, Texas, USA

Bar-Hillel, Y. 1960. The Present Status of Automatic Translation of Languages, *Advances in Computer* 1: 91-163.

Bentivogli, L., A. Bisazza, M. Cettolo, e M. Federico. 2016. Neural versus Phrase-Based Machine Translation Quality: a Case Study.

Castilho, S., J. Moorkens, F. Gaspari, I. Calixto, J. Tinsley, e A. Way. 2017. Is Neural Machine Translation the New State of the Art?. *The Prague Bulletin of Mathematical Linguistics* 108 (1): 109-20.

Castilho, S., S. Doherty, F. Gaspari, e J. Moorkens. 2018. Approaches to Human and Machine Translation Quality Assessment. In *Translation Quality Assessment: From Principles to Practice*. 2018, 9-38.

Castilho, S., J. Moorkens, F. Gaspari, M. Popović, e A. Toral. 2019. Editors' foreword to the special issue on human factors in neural machine translation. *Machine Translation*

Dicionário Terminológico para consulta em linha. Conectores Discursivos. <http://dt.dge.mec.pt/index.php?id=n442>. Consultado a 25 de março de 2021

García, M. 2009. La competencia conversacional de estudiantes de español como lengua extranjera: análisis y propuesta didáctica. Tese de doutoramento. Alcalá de Henares: Universidade de Alcalá.

Kenny, D. 2018. "Machine translation". Em *The Routledge handbook of translation and philosophy*, 428-45. New York: Routledge.

Hutchins, J., & Somers, H. 1992. *An Introduction to Machine Translation*. London: Academic Press Limited.

Hutchins, W.J. (ed) 2000. *Early Years in Machine Translation. Memoirs and Biographies of Pioneers*, 2. Amsterdam/Philadelphia: John Benjamins Publishing.

Hutchins, W.J. 2006. The First Public Demonstration of Machine Translation: The Georgetown-IBM System, 7th January 1954.

Lommel, A. 2018. Metrics for Translation Quality Assessment: A Case for Standardizing Error Typologies. Em *Translation Quality Assessment: From Principles to Practice*. 2018, 109-127.

Marcuschi, L. A. 2003. *Análise da Conversação*. 5^a Edição. São Paulo: Ática.

Martín, E. 2000. Conversadores competentes. *DidactiRed*.(18/12/2000).

Disponível em: http://cvc.cervantes.es/aula/didactired/anteriores/diciembre_00/18122000.htm. Consultado a 25 de março de 2021.

Moreno Fernández, F. 2002: *Producción, expresión e interacción oral*. Madrid: Arco Libros.

“Multidimensional Quality Metrics (MQM) Definition” – MQM Core. <http://www.qt21.eu/mqm-definition/definition-2015-06-16.html#mqm-core>. Consultado a 23 de janeiro de 2021

“Multidimensional Quality Metrics (MQM) Definition” – Scoring (Non normative). <http://www.qt21.eu/mqm-definition/definition-2015-06-16.html#scoring>. Consultado a 23 de janeiro de 2021

Pimentel, A., 2012. Os marcadores conversacionais no ensino de Português Língua Estrangeira: um estudo de caso.

“QT21 - Introduction”. QT21 - Quality Translation 21. 2019. <http://www.qt21.eu/?target=Introduction>. Consultado a 23 de janeiro de 2021

“Quality Metrics”. QT21 - Quality Translation 21 <http://www.qt21.eu/quality-metrics/>. Consultado a 23 de janeiro de 2021

Sacks, H., Schegloff, E. & Jefferson, G. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50 (4), 696-73.

Santos, N. 2019. Reflexão sobre aspetos de variação linguística na tradução de Sula, de Toni Morrison.

Somers, H.L. 1993. Current research in Machine Translation. *Machine Translation* 7, 231–246.

Van Dijk, T. A. 1989. *La ciencia del texto: un enfoque interdisciplinario*. Barcelona: Edições Paidós.