

Jorge Manuel Botelho Fonseca

2º Ciclo de Estudos em Linguística

**“GORDURA É FORMOSURA” OU “MAGREZA É BELEZA” NA TRADUÇÃO
AUTOMÁTICA ESTATÍSTICA?**
Estudo de caso de tradução de textos clínicos de aplicação informática de saúde com
o Moses for Mere Mortals e Google Tradutor

2014

Orientador: Professora Doutora Belinda Maia

Classificação:
Ciclo de estudos:

Dissertação/relatório/Projeto/IPP:

Versão definitiva

Agradecimentos

A elaboração de uma tese nunca é fácil e não fora a contribuição, direta ou indireta, de uma série de pessoas, não teria sido possível levar a cabo este projeto. É a essas pessoas, conscientes ou não da sua importância para a realização deste trabalho, que gostaria de deixar o meu sincero agradecimento:

- À **Professora Doutora Belinda Maia**, pela orientação e espírito ativo de procura constante de novos caminhos, mas principalmente pela disponibilidade e afabilidade com que sempre me recebeu;
- À **ALERT** por ter facilitado a minha frequência no mestrado em Linguística, pelo apoio logístico prestado e pela autorização para utilizar conteúdo proprietário;
- Aos meus **colegas de trabalho** na ALERT, em especial à Patrícia Ferreira pela revisão da tradução dos textos clínicos;
- Aos **colegas do Departamento de Línguas e Comunicação** da ALERT pelo incentivo e apoio constante, em especial à Sofia Fernandes pelas dicas preciosas sobre o texto da tese;
- Ao **Félix do Carmo e Luís Trigo**, que tão bem me acolheram no seu grupo de partilha de conhecimentos de doutoramento e cuja orientação, apoio e ensinamentos foram tão importantes para este projeto que, sem eles, nem teria sequer começado;
- À **Suzana Simões** que, sem me conhecer, partilhou abertamente todo o seu conhecimento sobre o Moses for Mere Mortals e que lançou definitivamente este projeto;
- Ao **Dr. Hilário Fontes**, da Direção Geral da Tradução da Comissão Europeia, pela imensa paciência que teve para ler e responder de forma tão pronta e detalhada a todos os meus *e-mails*. A forma tão espontânea, aberta e genuína com que partilhou todos os seus vastos conhecimentos na área da Linguística Computacional, nomeadamente ao nível do Moses for Mere Mortals, constituiu um exemplo para mim. Sem a sua ajuda este projeto não teria definitivamente chegado a bom porto;
- A toda a minha **família** e **amigos** pela compreensão pelas minhas muitas ausências e pelo apoio constante e incondicional;

- Ao meu **avô** que assistiu ao início deste projeto, mas que já não pôde testemunhar a sua conclusão, pelos valores da humildade e trabalho que incutiu em mim através do seu exemplo de vida;
- Aos meus **sogros**, cujo apoio doméstico e moral foi tão ou mais importante do que qualquer outro contributo intelectual para este projeto;
- Aos meus **pais** por todo o apoio de todos os géneros que me deram ao longo deste projeto e ao longo de toda a minha vida, eles que são o exemplo humano e profissional que procuro seguir;
- À minha **esposa**, Cláudia, e à minha **filha**, Ema, que apareceu para dar brilho às nossas vidas a meio deste projeto. O vosso apoio e incentivo constantes foram o mais importante de tudo e sem a vossa presença nenhuma conquista tem valor.

Resumo

Atualmente os sistemas de tradução automática estatística (TAE) constituem o paradigma dominante da investigação na área da tradução automática (TA) (Specia, 2013: 74). Estes sistemas analisam estatisticamente *corpora* paralelos (grandes quantidades de textos traduzidos e alinhados) para desenvolver modelos que permitam traduzir textos num par de línguas sem utilizar regras gramaticais.

No âmbito deste projeto procurámos saber se o adágio popular “gordura é formosura” também se aplicaria a sistemas de TAE, ou seja, se quantos mais dados o *corpus* de treino tiver melhores serão os resultados na tradução de textos de um domínio específico. Ou se pelo contrário “a magreza é beleza”, isto é, se a “qualidade” (entendida como pertinência dos dados), independentemente da quantidade, produz melhores resultados.

Para tal procedemos à tradução de inglês para português de dois documentos com textos clínicos utilizados no *software* clínico ALERT®. Ambos os textos foram traduzidos por dois sistemas de TAE: a versão simplificada do sistema livre e de código aberto Moses – o Moses-for-Mere-Mortals (MMM) – e a aplicação informática em linha Google Tradutor. Foram criados três motores MMM e treinados com três tipos de *corpora*. O primeiro *corpus* (ALL), que serviu de base aos restantes, foi composto por memórias de tradução da ALERT; o segundo, além de conter o *corpus* base, incluiu ainda o *corpus* da *European Medicines Agency* (EMA); e o terceiro, além de conter os outros dois *corpora*, incluiu ainda o *corpus* Europarl, que consiste em transcrições das sessões do Parlamento Europeu. As traduções foram avaliadas de forma automática e manual. Para a avaliação automática foi utilizada a métrica BLEU. Para a avaliação manual foi criado um sistema de categorização dos problemas por tipo e nível de gravidade.

Os resultados apontam para que nem a “gordura é formosura” nem a “magreza é beleza”, mas que “a virtude se encontra no meio”, *i.e.* não é exclusivamente a quantidade ou a “qualidade” do *corpus* de treino que produz os melhores resultados num sistema de TAE, mas deve ser um equilíbrio entre ambas.

Palavras-chave: tradução automática estatística, *corpus*, quantidade, qualidade, textos clínicos

Índice

Agradecimentos	i
Resumo	iii
Índice de figuras	vi
Índice de tabelas	ix
Lista de acrónimos.....	x
1. Introdução	1
1.1. Multilinguismo e Informática	1
1.2. Informática na Saúde e na Tradução.....	2
1.3. Apresentação do ALERT®.....	3
1.4. Estudo de Caso: Tradução de Textos Clínicos para Aplicação Informática de Saúde.....	4
2. Tradução Automática	5
2.1. Marcos na história da TA.....	5
2.1.1. A génese (1945-55)	5
2.1.2. O entusiasmo da comunidade científica (1956-65).....	7
2.1.3. Da desilusão à prática (1966-75).....	9
2.1.4. Implementação e primeiros frutos (1976-85).....	10
2.1.5. Novas necessidades: localização e ferramentas de apoio à tradução (1986-95)	11
2.1.6. <i>Corpora</i> e Internet lançam definitivamente sistemas TA (desde 1995).....	13
2.2. Abordagens à Tradução Automática.....	13
2.2.1. Tradução baseada em regras	14
2.2.2. Tradução baseada em <i>corpora</i>	18
2.2.3. Sistemas de tradução híbridos	22
2.3. Avaliação da tradução automática	22
2.3.1. Avaliação humana	23
2.3.2. Avaliação automática	24
2.4. Aplicações de sistemas TA	26
3. Sistemas de Tradução Automática Estatística	31
3.1. A Teoria do Canal Ruidoso	32
3.2. Modelo de tradução	35
3.2.1 Modelos baseados na palavra.....	35
3.2.2. Modelos de tradução baseados em <i>phrases</i>	40

3.3. Decodificador	43
3.4. Modelo de língua	49
3.5. A adição de conhecimento linguístico	54
4. Métodos e Materiais	59
4.1. Sistemas de TA testados	59
4.1.1. Moses for Mere Mortals	60
4.1.2. Google Tradutor	67
4.2. Textos clínicos ALERT®	69
4.3. Instrumentos de avaliação dos resultados da TA	71
4.3.1 Avaliação automática	72
4.3.2. Avaliação manual	74
5. Resultados	78
5.1. Resultados da avaliação manual	78
5.1.1. Resultados da tradução de “Instruções de Alta”	78
5.1.2. Resultados da tradução de “Textos Mais Frequentes”	84
5.1.3 Comparativo entre resultados da tradução de “Instruções de Alta” e “Textos Mais Frequentes”	91
5.2. Resultados da avaliação automática	98
6. Discussão de Resultados	101
7. Considerações Finais	106
Bibliografia	114
Referências <i>web</i>	118
Anexos	120

Índice de figuras

Figura 1 – Pirâmide de Vauquois que ilustra os diferentes níveis de complexidade dos modelos de tradução baseados em regras gramaticais.....	14
Figura 2 – Representação do processo de tradução direta (Hutchins, 2011: 7).....	15
Figura 3 – Representação do processo de tradução por transferência (Hutchins, 2007: 15).....	16
Figura 4 – Representação do processo de tradução por interlíngua (Hutchins, 2007: 18).	18
Figura 5 – Representação do processo de tradução baseada em exemplos (Hutchins, 2011: 10).....	20
Figura 6 – Representação do processo de tradução estatística (Hutchins, 2011: 9).	21
Figura 7 – Exemplo da aplicação da teoria de Shannon à tradução (Specia, 2010: 7)...	32
Figura 8 - Teoria do canal ruidoso de Shannon com base em Specia (2010: 7).	33
Figura 9 – Esquema da interação entre os componentes de um sistema de TAE com base em Knight e Koehn (2003: 4-5).....	34
Figura 10 – Todas as palavras devem estar alinhadas umas com as outras. No segundo exemplo, existe uma inconsistência porque há um ponto de alinhamento que se encontra fora do par de <i>phrases</i> (Koehn, 2010: 132).	41
Figura 11 – Tipos de orientação de <i>phrases</i> de acordo com o modelo de reordenação lexicalizada: (m) monótona, (s) <i>swap</i> (troca) e (d) descontínua (Koehn, 2010: 143). ...	43
Figura 12 – Diagrama exemplificativo da expansão de hipóteses para a tradução da frase "Mary did not slap the green witch." (Knight e Koehn, 2003: 78).....	46
Figura 13 - Exemplos de recombinações de Koehn (2010: 161)	46
Figura 14 – Hipóteses com o mesmo número de palavras são empilhadas num mesmo grupo. Uma hipótese é expandida ao ser aplicada a outra hipótese, dando origem a uma nova hipótese que é colocada numa nova pilha. (Koehn, 2010: 164)	47
Figura 15 – Diagrama da integração dos diferentes componentes do MMM (Machado <i>et al.</i> : 1).....	61
Figura 16 – <i>Scripts</i> na pasta Moses-for-Mere-Mortals-1.16.	62
Figura 17 – Área de configuração de parâmetros do <i>script</i> "create-1.36".	63

Figura 18 – Área desaconselhada a utilizadores menos avançados e que é constituída pela linha de código que executa as ações definidas pelos parâmetros selecionados pelo utilizador.....	63
Figura 19 – Aplicações Extract_TMX_Corpus_1.043 (esquerda) e Moses2TMX_1.032 (direita).	64
Figura 20 – Página do tradutor automático instantâneo Google Tradutor.	68
Figura 21 – Exemplares de textos pré-definidos utilizados pelo médico na atribuição de alta no ALERT®.	69
Figura 22 – <i>Script</i> “score-0.85” do MMM.	72
Figura 23 – Exemplo de tabela usada para a análise, catalogação e classificação dos problemas detetados em cada segmento.	76
Figura 24 – Exemplo de duas categorias de problemas detetados numa única palavra.	76
Figura 25 – Exemplo de atribuição de nível de gravidade quando a mesma categoria de problema ocorre mais do que uma vez.	77
Figura 26 – Número de ocorrências totais de problemas nas traduções de “Instruções de alta” realizadas pelo MMM e Google Tradutor.	78
Figura 27 – Pontuação média por segmento nas traduções de “Instruções de alta” realizadas pelo MMM e Google Tradutor.	79
Figura 28 – Percentagem de ocorrências por nível de impacto nas traduções de “Instruções de alta” realizadas pelo MMM e Google Tradutor.....	79
Figura 29 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de “Instruções de alta” realizada pelo MMM com o corpus ALL.	80
Figura 30 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de “Instruções de alta” realizada pelo MMM com o corpus EMEA.	81
Figura 31 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de “Instruções de alta” realizada pelo MMM com o corpus Europarl.....	82
Figura 32 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de “Instruções de alta” realizada pelo Google Tradutor.	83
Figura 33 – Ocorrências por categoria de problema na tradução de “Instruções de alta” realizada pelo MMM e Google Tradutor.....	84
Figura 34 – Número de ocorrências totais de problemas nas traduções de “Textos mais frequentes” realizadas pelo MMM e Google Tradutor.....	85

Figura 35 – Pontuação média por segmento das traduções de "Textos mais frequentes" realizadas pelo MMM e Google Tradutor.	85
Figura 36 – Percentagem de ocorrências por nível de impacto na tradução de “Textos mais frequentes” realizada pelo MMM e Google Tradutor.	86
Figura 37 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes " realizada pelo MMM com o <i>corpus</i> ALL.	87
Figura 38 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes" realizada pelo MMM com o <i>corpus</i> EMEA.	88
Figura 39 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes " realizada pelo MMM com o <i>corpus</i> Europarl...	89
Figura 40 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes " realizada pelo Google.	90
Figura 41 – Ocorrências por categoria de problema nas traduções de "Textos mais frequentes" realizadas pelo MMM e Google Tradutor.	91
Figura 42 – Número de ocorrências totais de problemas dos textos traduzidos pelo MMM e Google Tradutor.	92
Figura 43 – Número de ocorrências totais de problemas (sem contabilizar “M/M”) dos textos traduzidos pelo MMM e Google Tradutor.	93
Figura 44 – Pontuação média por segmento nos textos traduzidos pelo MMM e Google Tradutor.	94
Figura 45 – Pontuação média por segmento (sem contabilizar “M/M”) nos textos traduzidos pelo MMM e Google Tradutor.	95
Figura 46 – Distribuição de níveis de impacto por texto traduzido pelo MMM e Google.	96
Figura 47 – Distribuição de níveis de impacto por texto traduzido pelo MMM e Google (sem contabilizar “M/M”).	97
Figura 48 – Pontuação BLEU por sistema e por texto.	99

Índice de tabelas

Tabela 1 – Categorias de problemas detetados na tradução.	74
Tabela 2 – Níveis de gravidade/impacto de problemas detetados na tradução.	75
Tabela 3 – Ocorrências totais por categorias de problemas.	98

Lista de acrónimos

Acrónimo	Palavra por extenso
MMM	Moses for Mere Mortals
M/M	Maiúsculas/minúsculas
TAE	Tradução automática estatística
TA	Tradução automática
TP	Texto de partida
TC	Texto de chegada
LP	Língua de partida
LC	Língua de chegada
IA	“Instruções de Alta”
TMF	“Textos Mais Frequentes”
Cf.	Confrontar

1. Introdução

“e o Senhor disse: Eis que o povo é um, e todos têm uma mesma língua; e isto é o que começam a fazer; e, agora, não haverá restrição para tudo o que eles intentarem fazer.”

Livro do Génesis, 11: 6

1.1. Multilinguismo e Informática

A consciência de que a variedade de línguas é um entrave ao desenvolvimento pleno da humanidade está presente desde tempos imemoráveis e expressa-se de forma vincada na passagem do Livro do Génesis que relata o episódio da Torre de Babel. Reza a história bíblica que Deus castigou os Homens separando as línguas por estes pretenderem construir uma torre que tocasse o céu e perpetuasse a sua fama. Seria desde essa altura que os homens procuram recuperar a capacidade perdida de comunicar livremente, *i.e.* sem constrangimentos linguísticos, entre todos os povos da Terra.

A imensa potencialidade dos computadores, desenvolvidos especialmente a partir de meados do século passado, veio dar nova força e esperança à demanda de recuperar a capacidade de comunicação universal. Desde que Warren Weaver, na década de 40, lançou as bases para a utilização dos computadores como “descodificadores” das línguas, que as esperanças de quebrar as barreiras linguísticas se têm renovado, como nunca, nos últimos tempos. Isso mesmo expressa Nicholas Ostler, na sua obra *The Last Lingua Franca* em que reforça o papel da tecnologia na superação dos obstáculos linguísticos:

“But just as the print revolution changed the ground rules of communication among Europeans in the sixteenth century, so modern electronic technology is set to change the ancient need for a single lingua franca. When electronics removes the requirement for a human intermediary to interpret or translate, the frustrations of the language barrier may be overcome without any universal shared medium beyond compatible software.”

Nicholas Ostler, 2010: xix

Nicholas Ostler defende que o papel de língua franca que o inglês parece assumir atualmente deixará de existir a longo prazo e que não mais existirão outras línguas francas, uma vez que o futuro, em grande parte impulsionado pela tecnologia, será do multilinguismo.

A crença de Ostler nas capacidades da tecnologia informática apresenta muitas semelhanças com as crenças iniciais acerca das potencialidades da tradução automática (TA), na década de 50 do século passado¹. Contudo, apesar de essas crenças, até ao momento, não se terem materializado por completo, e após um grande período de descrédito e desalento na TA², a verdade é que a mesma tem vindo a evoluir – mesmo que não seja ao ritmo desejado ou inicialmente estimado – e se tem instalado cada vez mais no nosso quotidiano, devido em grande parte à expansão da Internet. Embora a qualidade deste tipo de tradução não tenha ainda atingido o nível da tradução humana, tem vindo a melhorar gradualmente, mas, mais importante do que isso, tem permitido a utilizadores de todo o mundo aceder a informação a que não teriam acesso de outra forma³. Daí que, crendo-se ou não na capacidade da tecnologia informática para contornar o castigo divino relatado no Génesis, a verdade é que esta tem vindo cada vez mais a ajudar os Homens a ultrapassar as barreiras linguísticas e a facilitar a comunicação entre povos de todo o mundo.

1.2. Informática na Saúde e na Tradução

Na era do conhecimento, impulsionada pelo desenvolvimento da Internet, e da Web 2.0, onde cada utilizador contribui ativamente com informação na rede global, as questões linguísticas assumem uma importância crescente. Prova disso, tem sido o desenvolvimento que a Linguística Computacional ou Processamento de Linguagem Natural (PLN) tem conhecido. A área de abrangência da PLN tem vindo a expandir-se e quase se poderá dizer que não haverá área da atividade humana em que a PLN não tenha, ou possa vir a ter, influência⁴.

¹ Vide o subcapítulo 2.1.2. “O entusiasmo da comunidade científica (1956-65)”, pp. 16-17

² Vide o subcapítulo 2.1.3. “Da desilusão à prática (1966-75)”, pp. 17-18.

³ É o caso, por exemplo do Google Tradutor, que permite aos utilizadores da Internet aceder a conteúdos eletrónicos numa panóplia de línguas que seria impensável até há pouco tempo.

⁴ Cf. Bose, 2004

A investigação da TA, enquanto uma das áreas de interesse do PLN ou da linguística computacional, foi sendo, desde o seu início, orientada para domínios muito restritos, visto que a sua utilidade servia essencialmente interesses económicos e políticos⁵. Atualmente, apesar de a TA ser mais conhecida do público em geral na tradução de textos do domínio geral, a sua maior utilidade continua a ser para domínios específicos do conhecimento.

Uma das áreas que pode, direta ou indiretamente, beneficiar dos desenvolvimentos da TA é o setor da saúde. Sendo uma área onde é produzida cada vez mais informação e para a qual a informação é cada vez mais fundamental, a TA pode ser uma ferramenta importante no apoio à disseminação e partilha dessa mesma informação. A TA tanto pode ser útil no apoio à tradução do crescente número de artigos e documentos de prática clínica, como pode ainda integrar outras ferramentas de apoio à atividade clínica. Nesse âmbito, pode, por exemplo, integrar ferramentas de exploração de texto (*text mining*) em *software* clínico ou sistemas multilingues de pergunta-resposta⁶. É, no entanto, de destacar que, além das enormes potencialidades que a TA apresenta neste setor de atividade, a utilização da mesma apresenta também uma dose de perigo na mesma proporção, uma vez que os seus resultados não são (pelo menos atualmente) isentos de erro, o que, numa área tão sensível como a da saúde, obriga a que este tipo de ferramentas seja usado com a devida cautela.

A temática deste trabalho insere-se na importância que a TA pode assumir no apoio à tradução de conteúdos relacionados com as ciências da vida, nomeadamente no âmbito da atividade do Departamento de Línguas e Comunicação da empresa de *software* clínico ALERT Life Sciences Computing.

1.3. Apresentação do ALERT®

A ALERT Life Sciences Computing S.A. é uma empresa multinacional portuguesa que opera na área da informática clínica e que tem como missão “melhorar a saúde e prolongar a vida”⁷. Fundada em 1999, dedica-se à criação de uma solução integrada para a total informatização dos cuidados de saúde em países inteiros, organizações de

⁵ Vide subcapítulo 2.1.1. “A génese (1945-55)”, pp 14-16.

⁶ Cf. subcapítulo 2.4 “Aplicações de sistemas de TA”, pp. 26-31.

⁷ Vide <http://www.alert-online.com/pt/mission-values>.

manutenção de saúde (*Health Management Organizations*), hospitais e centros de saúde, passando pelos processos individuais de cada paciente. Os produtos da ALERT abrangem desde o local de prestação dos cuidados de saúde (através de um Processo Clínico Eletrónico para profissionais, destinado a ser utilizado em (1) hospitais, (2) centros de saúde, (3) clínicas) ao domicílio do paciente (através do Processo Clínico Eletrónico Pessoal para registo de dados e repositório de informação clínica além fronteiras e entre diferentes sistemas de soluções clínicas). Os produtos ALERT® incluem ainda *software* de referenciação e agendamento, assim como aplicações para planeamento de cuidados de saúde. Além disso, a ALERT disponibiliza ainda bases de conhecimento pessoal para faculdades de medicina, estudantes e profissionais de saúde. O ALERT® é atualmente a solução clínica mais internacional existente, tendo sido já adotado em 13 países (Portugal, Espanha, França, Suíça, Reino Unido, Itália, EUA, Brasil, México, Chile, Guatemala, Angola e Kuwait), e encontra-se disponível em 6 línguas (português, inglês, espanhol, francês, italiano e neerlandês).⁸

1.4. Estudo de Caso: Tradução de Textos Clínicos para Aplicação Informática de Saúde

Nesse âmbito, decidimos realizar um estudo de caso de tradução de textos clínicos utilizados na aplicação informática de saúde ALERT®. Desafiando aquilo que Axelrod, Xiaodong, e Jianfeng (2011) denominam de “sabedoria convencional”, ou seja, a crença de que “mais dados é melhor; quanto maior for um *corpus* de treino, mais exato será o modelo”⁹, pretendemos avaliar se um sistema de base meramente estatística – neste caso, o Moses for Mere Mortals – treinado com *corpora* bilingues mais pequenos, mas com predominância de dados específicos do ALERT® (*in-domain*), dados do domínio da saúde em geral (*domain*) e dados fora do domínio da saúde (*out-of-domain*) é capaz de produzir melhores resultados do que um sistema com um *corpus* consideravelmente maior, mas sem orientação a qualquer domínio específico – neste caso, o Google Tradutor.

⁸ Para mais informações acerca da ALERT e seus produtos visitar: <http://www.alert-online.com/>

⁹ Tradução nossa.

2. Tradução Automática

Neste capítulo iremos abordar vários aspetos da tradução automática (TA). Começaremos por traçar a sua origem e percorrer os momentos mais marcantes, para depois nos determos nos diferentes modelos utilizados na investigação da TA. Além disso, iremos ainda mencionar a importância da avaliação humana e automática para a orientação da investigação em TA e, finalmente, destacar as principais utilidades da TA.

2.1. Marcos na história da TA

Com o advento dos computadores e o desenvolvimento destes, principalmente durante a 2ª Guerra Mundial, para a codificação e decodificação de mensagens, ganha cada vez mais forma a ideia de utilizar este tipo de máquinas para a tradução das línguas naturais. Ao longo da sua história, a tradução automática tem beneficiado mais dos desenvolvimentos tecnológicos do que de influências de correntes linguísticas para a sua evolução. Embora a qualidade dos sistemas de TA não tenha conseguido acompanhar o ritmo evolutivo da tecnologia informática que os suporta, a utilidade destes sistemas tem extravasado o espaço confinado das universidades e das empresas onde nasceu e atualmente tem cada vez maior presença no quotidiano de cada cidadão num contexto de uma sociedade crescentemente informatizada e globalizada. Como tal, convém recuar cerca de meia década e recordar o percurso dos sistemas de TA até aos dias de hoje.

2.1.1. A génese (1945-55)

Em meados da década de 40, David Booth, um cientista computacional britânico, e Warren Weaver, diretor da Rockefeller Foundation¹⁰ discutem ideias acerca das potencialidades dos recém-criados computadores para a tradução de línguas naturais. Incentivado por colegas da fundação, em 1949, Warren Weaver explana as suas ideias num memorando que ficou célebre por estabelecer as linhas orientadoras e os objetivos de investigação (mais do que abordar questões técnicas) de uma nova área de aplicação

¹⁰ A Rockefeller Foundation é uma instituição privada norte-americana de cariz humanitário que atua não só no território americano, como em outras partes do mundo.

dos computadores: a tradução automática. Este documento não só chamou a atenção para uma possível nova utilidade dos computadores, como indicou as potencialidades da estatística na análise linguística e abordou possíveis bases lógicas das línguas e a existência de elementos universais entre estas (Hutchins, 1986: 7). Embora as suas ideias na época tenham sido recebidas com algum ceticismo por parte de linguistas¹¹, é hoje reconhecido que as mesmas deram um contributo e estímulo fundamental para a investigação em TA nos EUA (Hutchins, 2006: 2). Nos dois anos seguintes, assistiu-se ao aparecimento de várias unidades de investigação na área da TA, nomeadamente no Massachusetts Institute of Technology (MIT), na Universidade de Washington, em Seattle, na Universidade de Califórnia, em Los Angeles (UCLA), no National Bureau of Standards¹², na RAND Corporation¹³ e em vários outros locais nos EUA (Hutchins, 1986: 8). Em 1951, Yehoshua Bar-Hillel tornou-se o primeiro investigador totalmente dedicado a esta nova área. A sua missão passava por avaliar as possibilidades da TA e planejar a investigação científica nesse campo para o futuro (*ibidem*: 10). No relatório sobre a situação da investigação em TA na época, publicado no final do mesmo ano, Bar-Hillel aborda as problemáticas que irão dominar as discussões de TA nos anos seguintes: a viabilidade de uma tradução totalmente automática, o papel da pós-edição, os objetivos da análise sintática, a importância da informação estatística, a possibilidade de obter uma gramática universal, os fundamentos lógicos das línguas, e os vocabulários restritos (Hutchins, 1986: 10). Apesar de não acreditar na viabilidade do desenvolvimento de sistemas capazes de realizar a tradução de forma autónoma e totalmente automática, produzindo resultados de elevada qualidade num futuro próximo, Bar-Hillel considera mais realista uma meta que inclua a presença humana no processo de tradução de forma a dar conta dos aspetos que, de outra maneira, seria impossível às máquinas processar (como, por exemplo, ambiguidades semânticas e aspetos estilísticos dos textos) (*ibidem*). Bar-Hillel considera ainda a possibilidade de se criar gramáticas gerais de línguas artificiais ou gramáticas de transferência em que “a gramática de uma língua é expressa em categorias adequadas para outra língua”¹⁴. Além disso, propõe que a TA seja utilizada em situações onde ocorra vocabulário restrito (por

¹¹ Cf. Weaver, 1949: 5 e Hutchins, 1986: 7

¹² Atualmente a National Institute of Standards and Technology (NIST) é uma agência governamental estadunidense que se dedica a promover a inovação e a competitividade industrial nos EUA.

¹³ A RAND Corporation é uma instituição sem fins lucrativos estadunidense que se dedica a apoiar o desenvolvimento de políticas e a tomada de decisões através da investigação.

¹⁴ Tradução nossa de Bar-Hillel, 1951 *apud* Hutchins, 1986: 10.

exemplo, aeronáutica ou meteorologia), línguas regularizadas (como o inglês básico¹⁵) ou línguas auxiliares (como o esperanto ou interlíngua) (*ibidem*).

Em 1952 Bar-Hillel organiza no MIT a primeira conferência inteiramente dedicada à TA. Os principais temas de debate foram as questões de pré e pós-edição, línguas controladas, domínios restritos, análise sintática, *hardware* informático, programação e financiamento (Hutchins, 2007: 4).

Um dos resultados mais importantes deste encontro foi a criação de uma equipa de investigação na Universidade de Georgetown liderada por Leon Dostert com a missão de criar um sistema piloto capaz de demonstrar na prática as potencialidades da TA. De certa forma, foi o primeiro passo na transição de uma abordagem teórica para uma abordagem empírica na investigação da TA (Hutchins, 1986: 13).

Em parceria com a IBM, a Universidade de Georgetown organizou uma sessão pública para demonstrar as potencialidades do seu sistema de TA na tradução de um *corpus* limitado e cuidadosamente selecionado de frases em russo¹⁶. O acontecimento revestiu-se da maior importância, não só pela dimensão que adquiriu junto do público em geral (e da própria comunidade científica) – que permitiu encarar a TA como um objetivo viável e conduziu a um forte investimento por parte do governo -, mas também porque foi a primeira demonstração real de TA num computador (experiências anteriores limitaram-se a simulações feitas à mão ou em máquinas de cartões perfurados) (*ibidem*: 12-13). Além de ter estimulado a investigação nesta área em larga escala por todo o território dos EUA, esta demonstração inspirou também a iniciação de projetos semelhantes noutros pontos do mundo (nomeadamente na Europa e URSS).

2.1.2. O entusiasmo da comunidade científica (1956-65)

Influenciadas pela publicidade que a demonstração de Georgetown teve tanto nos EUA como no resto do mundo, surgem várias equipas dedicadas à investigação não só no território dos EUA, como também na Europa, URSS e Ásia (nomeadamente na China e

¹⁵ O inglês básico é uma língua controlada que tem por base o inglês. Criada por Charles Kay Ogden, linguista e filósofo, na sua obra *Basic English: A General Introduction with Rules and Grammar* (1930), surge como língua auxiliar internacional e como forma de apoio ao ensino do inglês como segunda língua. (http://en.wikipedia.org/wiki/Basic_English acedido em fevereiro de 2014).

¹⁶ A escolha do par de línguas não era inocente em relação ao momento político que se vivia na época, com o início do arrefecimento das relações entre EUA e URSS após o final da II Guerra Mundial. Motivações políticas fizeram com que a investigação em TA na década de 50 fosse financiada pelos governos tanto dos EUA como da URSS (Frequently asked questions about Computational Linguistics s.d.).

Japão) (Hutchins, 2010: 2). Surgem cada vez mais publicações sobre a temática da TA um pouco por todo o mundo, assim como os encontros entre especialistas. A primeira conferência internacional surge em 1956, organizada pelo MIT, o que revela o interesse crescente que o tema vai suscitando um pouco por toda a parte (Hutchins, 1986: 13).

De acordo com Hutchins (ibidem), um aspeto que se começa a revelar, assim que começam a surgir vários grupos de estudo, é que as abordagens adotadas tendem a polarizar-se. Se, por um lado, há aqueles que colocam a ênfase nas problemáticas relacionadas com dicionários, lexicografia e semântica (por exemplo, Universidade Washington, UCLA e RAND), por outro, há aqueles que se debruçam com maior afinco nas questões sintáticas (Universidade de Georgetown e MIT). Da mesma forma, há aqueles que seguem uma abordagem teórica (MIT e Cambridge) e os que preferem uma abordagem prática (Georgetown e RAND). Finalmente, há aqueles que pretendem desenvolver sistemas operacionais para um futuro próximo (Washington e Georgetown, entre outros) e aqueles que têm o objetivo de longo prazo de produzir sistemas para traduções de elevada qualidade (Harvard e MIT).

Sendo uma área de investigação recente, que requeria uma tecnologia muito dispendiosa, Europa e União Soviética, ainda a recuperar dos danos da II Guerra Mundial, eram inevitavelmente forçados a concentrar o seu foco em questões de cariz principalmente teórico. Devido a motivações políticas e militares, o principal par de línguas utilizado na investigação desta área era o inglês e o russo, embora a política multilingue adotada pela República Soviética inspirasse a investigação num conjunto mais alargado de línguas do que em qualquer outro local (Hutchins, 2006: 377).

Em meados dos anos 60 existiam já vários grupos de investigação em TA espalhados por todo o mundo, embora muitos deles fossem de curta duração. Léon (1997), *apud* Hutchins (2006: 377), refere que a investigação em TA, durante este período, agregou a investigação de muitas outras áreas (como a linguística estruturalista, semiótica, semântica lógica, linguística matemática, linguística quantitativa e quase tudo o que hoje se designa por linguística computacional).

As perspetivas nesta época eram as melhores, mercê dos desenvolvimentos na área da informática e da linguística formal (especialmente ao nível da sintaxe) (Hutchins, 2010: 5). Esperava-se que fosse possível conseguir realizar traduções de boa qualidade de forma totalmente automática dentro de alguns anos. Contudo, a realidade veio a demonstrar que estas expectativas eram infundadas.

2.1.3. Da desilusão à prática (1966-75)

À medida que os problemas linguísticos se iam revelando e a investigação em TA parecia ter chegado a um beco, aumentava a desilusão em face das expectativas criadas (Hutchins, 2010: 5). Bar-Hillel (1960: 92-93) critica a crença de que a meta da TA deva ser a criação de sistemas capazes de produzir resultados de elevada qualidade, indistintos da produção humana, por considerar esta uma “ilusão”. De acordo com Bar-Hillel, a automatização completa do processo de tradução é incompatível com um resultado de elevada qualidade. Como tal, considera que apenas poderá haver dois caminhos a seguir: sacrificar a qualidade ou diminuir a independência da máquina na produção do resultado final da tradução, alterando o foco da investigação da tradução automática para a investigação de ferramentas automáticas de apoio à tradução (*ibidem*). Apesar de os primeiros sistemas terem demonstrado a utilidade da TA de baixa qualidade em variados contextos, continuava a crescer o sentimento de desilusão perante a incapacidade de a comunidade científica conseguir, por esta altura, apresentar sistemas de TA de elevada qualidade, conforme se esperava na década de 50¹⁷.

Em 1964, os patrocinadores da investigação de TA nos EUA (nomeadamente o exército e os serviços secretos estadunidenses) pedem à National Science Foundation que crie uma equipa – a Automatic Language Processing Advisory Committee (ALPAC) – para analisar a situação dos progressos feitos em TA e apontar caminhos para o futuro¹⁸. A conclusão do ALPAC, exposta num famoso relatório de 1966, é que a investigação na área da TA é lenta, cara e os sistemas produzidos apresentam resultados com qualidade inferior à tradução humana¹⁹. A adicionar a tudo isto, não havia ainda a perspetiva – na opinião dos autores do relatório – que este cenário mudasse a curto prazo²⁰. Como tal, o estudo desaconselha o investimento na investigação em TA e recomenda que as verbas sejam direcionadas para a investigação em linguística computacional, enquanto parte integrante da Linguística, e para o desenvolvimento de ferramentas eletrónicas de apoio à tradução²¹. O relatório ALPAC teve uma enorme influência junto das entidades que

¹⁷ Cf. Bar-Hillel, 1960: 92-93.

¹⁸ Cf. ALPAC, 1966.

¹⁹ Cf. ALPAC, 1966

²⁰ Cf. ALPAC, 1966: 25

²¹ Cf. ALPAC, 1966: 34

financiavam a investigação em TA, o que levou a que esta terminasse quase por completo nos EUA e, indiretamente, noutras partes do mundo (Hutchins, 2010: 6).

Apesar do rude golpe desferido por este relatório, a investigação em TA não cessou por completo, nem nos EUA nem no resto do mundo. Pelo contrário, permaneceu (com menos grupos, é certo) e explorou novas direções. Enquanto até aí a considerada primeira geração de sistemas de TA tinha seguido maioritariamente uma abordagem de tradução direta, a partir da segunda metade da década de 60 assistiu-se a uma exploração de métodos de tradução “indireta”, como as abordagens baseadas em transferência e em interlíngua (Hutchins, 2010: 6). Foi nesta e na década seguinte que tiveram início muitos projetos que viriam mais tarde a conhecer sucesso e reconhecimento não só por parte da comunidade científica como do público em geral.²²

2.1.4. Implementação e primeiros frutos (1976-85)

Apesar do revés sofrido em termos de financiamento após o relatório ALPAC, a investigação em TA manteve-se ativa, embora com menos recursos e entusiasmo mais refreado, e foi, paulatinamente, desenvolvendo sistemas com diferentes abordagens que iriam despontar e reavivar o interesse pela TA no final dos anos 70 e durante toda a década de 80.

Um dos mais significativos terá, porventura, sido o sistema Systran, cuja primeira versão foi instalada na Força Aérea dos EUA, em 1970, e mais tarde na Comissão das Comunidades Europeias, em 1976, além de em várias grandes empresas durante a década de 80.

Contrariamente à tendência verificada na década de 60 e início de 70 para a adoção de uma abordagem baseada em interlíngua, a partir da segunda metade da década de 70 verifica-se uma tendência para a tradução baseada em transferência (Hutchins, 2006: 378). Sistemas como o Ariane (desenvolvido pelo Groupe d’Étude pour la Traduction Automatique de Grenoble), o SUSY (desenvolvido na Alemanha, em Saarbrücken), o MU (desenvolvido pela Universidade de Kyoto, no Japão) e o Eurotra (sistema desenvolvido pela Comunidade Europeia) são alguns exemplos significativos de sistemas com esta abordagem à tradução (*ibidem*: 378-379).

²² Para mais informação *vide* Hutchins, 2010: 6-8 e 2006: 378-380.

Em meados da década de 80 ressurge o interesse por sistemas de interlíngua devido aos desenvolvimentos alcançados na área da inteligência artificial e da linguística cognitiva. Acreditava-se que os avanços ao nível da qualidade de produção de sistemas TA viriam da investigação em linguagem no contexto da inteligência artificial (*ibidem*: 379).

De acordo com Hutchins (*ibidem*), é também por esta altura que a investigação em TA conhece forte desenvolvimento no Japão, fomentada por investimentos de grandes empresas tecnológicas e do próprio governo, assim como noutras partes da Ásia. Na União Soviética o foco da investigação em TA está centrado na análise estatística, influenciado pelos desenvolvimentos realizados na área da estatística da fala pela Universidade de Leninegrado. Surgem vários projetos de melhoramento da qualidade da tradução com base no conhecimento no Japão, Europa e América do Norte.

É também na década de 80 que surge o controlo do vocabulário e sintaxe como forma de melhoramento da qualidade de TA. A empresa estadunidense Xerox cria restrições para controlar os textos que são utilizados como base da TA, de forma a facilitar a transferência e geração de texto na LC – um método que colheu vários seguidores nos anos subsequentes, chegando ao ponto de passarem a existir empresas dedicadas ao controlo da linguagem (*ibidem*; 380).

2.1.5. Novas necessidades: localização e ferramentas de apoio à tradução (1986-95)

Até finais da década de 80, os sistemas de TA eram construídos tendo por base um conjunto de regras linguísticas. A partir da década de 90 surgem novos métodos e estratégias baseados em grandes quantidades de texto. A possibilidade de utilizar *corpora* de grandes dimensões, mercê do avanço da tecnologia informática, reacende o interesse por abordagens estatísticas à tradução. É o regresso do empirismo, em oposição ao racionalismo dos métodos baseados em regras (Hutchins, 2006: 380).

Com o sucesso de técnicas estocásticas²³ no reconhecimento da fala, a IBM desenvolve um sistema baseado num modelo de tradução estatística – o Candide. A grande

²³ Na teoria da probabilidade, um processo estocástico é o contrário de um processo determinístico. Ou seja, num processo estocástico, o desenrolar do processo apresenta várias possibilidades indeterminadas, embora algumas possam ser mais prováveis do que outras (*vide* http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Stochastic_process.html).

inovação deste sistema consistia no facto de não utilizar quaisquer regras linguísticas na tradução, mas apenas estatística.

Para além dos métodos estatísticos, na década de 90 assiste-se também ao eclodir de um outro método baseado em *corpora*: a tradução baseada em exemplos. Apesar da primeira proposta neste sentido ter surgido em 1981 por Makoto Nagao, só no final dessa década é que se iniciaram experiências com esse modelo (Hutchins, 2006: 380). Ainda assim, a investigação de TA baseada em regras linguísticas prossegue, tanto em sistemas de transferência como de interlíngua (*ibidem*: 381). Iniciam-se experiências em sistemas denominados híbridos por conjugarem modelos baseados em regras com modelos baseados em *corpora* (Hutchins, 2010: 14).

Da investigação de TA surgiram, nos anos 90, áreas derivadas de investigação, como a tradução da fala, que combina o reconhecimento da fala com a interpretação e produção de fala (Hutchins, 2006: 381). É igualmente neste período que surge como grande foco de atenção da investigação a avaliação dos resultados da TA, como forma de orientar e qualificar os sistemas e esforços desenvolvidos (*ibidem*).

A utilização da TA expande-se fortemente a partir da década de 90. Empresas e entidades públicas, que normalmente necessitam da tradução de uma grande quantidade de conteúdos, principalmente de cariz técnico, começam cada vez mais a utilizar sistemas de TA, o que estimula o mercado neste setor (*ibidem*).

Hutchins (*ibidem*) refere que é também nesta década, com a disseminação dos computadores pessoais, que surgem as primeiras ferramentas comerciais de apoio à tradução que combinam várias ferramentas eletrónicas – processadores de texto multilingues, ferramentas de gestão de terminologia, memórias de tradução, etc. Os *mainframes*²⁴, essenciais à TA até finais da década de 80, são substituídos por computadores pessoais modernos, muito mais pequenos e baratos e com as mesmas capacidades de processamento. O público-alvo dos sistemas TA estende-se assim das grandes corporações para tradutores individuais. Reflexo do alargamento do mercado dos sistemas de TA é o número de empresas dedicadas a essa área que surgem durante os anos 80 e 90 (cf. Hutchins, 2010: 15-17).

²⁴ Computadores de grande porte capazes de realizar operações complexas (*vide* <http://pt.okpedia.com/mainframe>).

2.1.6. Corpora e Internet lançam definitivamente sistemas TA (desde 1995)

A disseminação da Internet a uma escala global, que se verificou a partir de meados da década de 90, e a importância crescente que este novo meio de comunicação adquire, exercem uma influência decisiva no desenvolvimento dos sistemas de TA. Serviços de TA passam a ser disponibilizados através da Internet para chegarem a um público ainda mais alargado. Apesar de a qualidade destes sistemas ainda ser relativamente fraca²⁵, estes serviços vêm dar resposta à crescente necessidade de traduções imediatas para fins de assimilação (Hutchins, 2010: 17). Além disso, numa era onde a informação circula livremente e de forma global, surge cada vez mais a necessidade de disponibilizar os conteúdos em várias línguas e quase em simultâneo. Perante este cenário, a TA surge como resposta evidente e imediata às novas necessidades de tradução de páginas Web, correio eletrónico e outras formas de comunicação através da Web. Além disso, os sistemas de apoio à tradução passam a integrar sistemas de TA, unindo-se, desta forma, os dois caminhos de investigação de TA cuja separação tinha sido formalmente recomendada no relatório ALPAC na década de 60. Igualmente importante foi a disponibilização de forma gratuita de serviços de TA *online* (Babelfish primeiro e Google Tradutor mais tarde, entre vários outros), tornando-se um produto disponível não apenas para profissionais, mas para o público em geral e para ser usado numa multiplicidade de situações e não apenas em textos técnicos (para os quais a TA tinha sido inicialmente concebida nas décadas de 50 e 60).

À medida que a TA vai sendo cada vez mais disponibilizada e conhecida junto de um público cada vez mais alargado, vão surgindo novas possibilidades de desenvolvimento em direções que tanto no passado como atualmente não conseguimos ainda perspetivar (Hutchins, 2010: 18).

2.2. Abordagens à Tradução Automática

A investigação da TA tem utilizado diversas abordagens na demanda da forma ideal de produzir traduções de elevada qualidade sem o apoio humano. Em termos de arquitetura computacional, podemos categorizar os sistemas de TA em sistemas baseados em

²⁵ A qualidade da tradução produzida por sistemas TA não consegue acompanhar os avanços verificados na tecnologia informática que os suporta (Hutchins, 2006: 381).

regras, em *corpora* e híbridos²⁶. No âmbito dos sistemas baseados em regras podemos distinguir três modelos essenciais: tradução direta, transferência e interlíngua²⁷. Em termos de sistemas de TA baseados em *corpora*, podemos distinguir sistemas baseados em exemplos (também conhecidos pela sua sigla inglesa EBMT – *Example-Based Machine Translation*) e sistemas de tradução automática estatística (daqui em diante TAE)²⁸. Os sistemas híbridos, por utilizarem tanto componentes baseados em regras como em *corpora* constituem uma abordagem à parte.²⁹

2.2.1. Tradução baseada em regras

A tradução baseada em regras, tal como o nome indica, representa o sistema de TA que utiliza regras gramaticais para a tradução. Este tipo de sistemas pode ser dividido em três categorias que representam, cada uma, uma abordagem diferente à tradução baseada em regras: tradução direta, tradução por transferência e tradução por interlíngua. O grau de complexidade de cada uma destas abordagens é representado no triângulo de Vauquois:

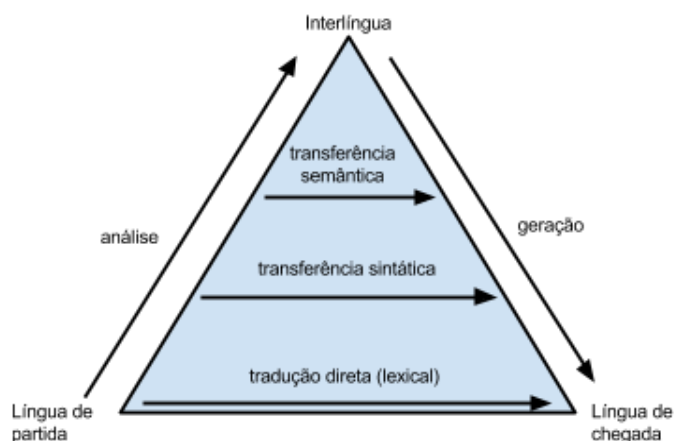


Figura 1 – Pirâmide de Vauquois que ilustra os diferentes níveis de complexidade dos modelos de tradução baseados em regras gramaticais.

²⁶ Conforme proposto por Chérargui, 2012: 165. Na ausência de uma classificação uniforme, adotada em geral pela investigação em TA, acerca das diferentes abordagens à TA (*vide*, por exemplo, Chérargui, 2012; Tripathi & Sarkhel, 2010; Euromatrix, 2007; Simões, 2009), assumimos a opção de fazer a nossa própria classificação com base em propostas de diferentes investigadores.

²⁷ Com base no proposto por Simões, 2009: 2.

²⁸ Com base no proposto por Simões, 2009: 2.

²⁹ Outras classes poderiam ser distinguidas e discriminadas, de acordo com diferentes autores (*vide*, por exemplo, Hutchins, 2003a, 2003b; Tripathi & Sarkhel, 2010; Chérargui, 2012; Simões, 2009).

De seguida é apresentada uma breve descrição de cada uma destas abordagens.

2.2.1.1. Tradução direta

A tradução direta (também designada “tradução binária”) é concebida para um par de línguas específico e é normalmente unidirecional (Hutchins, 2003b: 503). Nesta abordagem, a tradução é feita diretamente do texto de partida (TP) para o texto de chegada (TC) com o mínimo de análise sintática ou semântica para a resolução de ambiguidades na língua de partida (LP), identificação de expressões na língua de chegada (LC) e especificação da ordem das palavras na LC (*ibidem*). No modelo de tradução direta, as palavras da LP são despidas de inflexão morfológica de forma a obter a sua forma base (Chéragui, 2012: 163-164). Como tal, este tipo de sistemas é normalmente composto por um grande dicionário bilingue e um único módulo de análise e geração de texto (Hutchins, 2003b: 503). Depois de traduzidas, as palavras são organizadas de acordo com regras básicas de ordenação da LC (como, por exemplo, colocar o adjetivo a seguir ao nome ao traduzir de inglês para português) (Collins, (s.d.); 10)

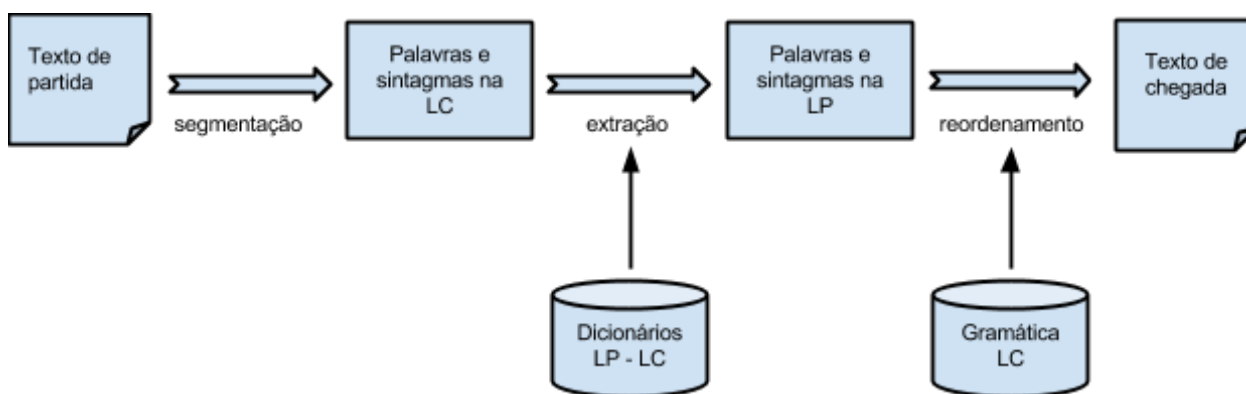


Figura 2 – Representação do processo de tradução direta (Hutchins, 2011: 7).

2.2.1.2. Tradução por transferência

Os sistemas com uma abordagem à tradução através de transferência são considerados a segunda geração de sistemas de TA (Hutchins, 2003b: 504). Esta abordagem é composta por três fases: 1) na primeira, denominada análise, o sentido do texto na LP é convertido em representações abstratas; 2) na segunda fase, designada transferência, as

representações abstratas da LP são convertidas em representações abstratas na LC; 3) e na terceira fase, conhecida por síntese, é gerado o TC (*ibidem*: 503-504).

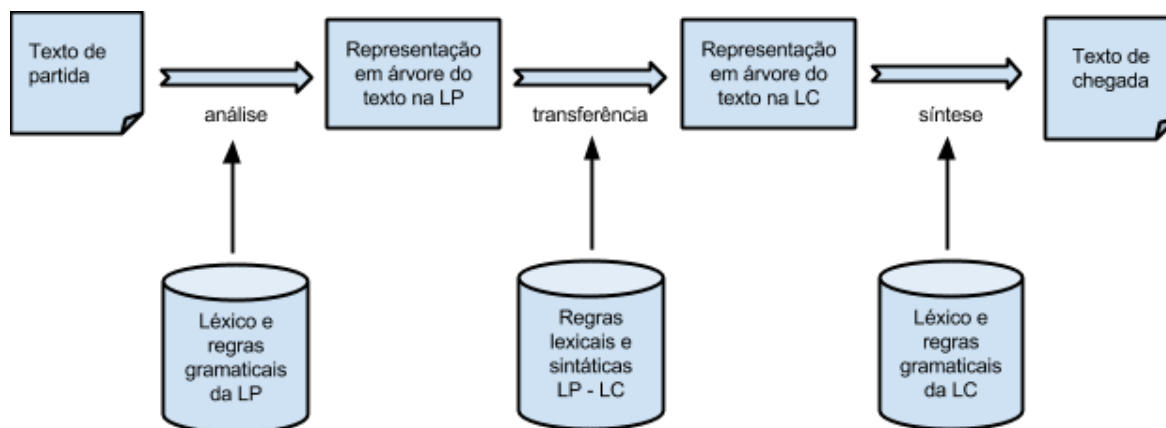


Figura 3 – Representação do processo de tradução por transferência (Hutchins, 2007: 15).

Nesta abordagem, os sistemas de TA são constituídos por dicionários da LP com informação morfológica, sintática e semântica, dicionários semelhantes na LC e um dicionário bilingue de transferência em que as formas base da LP são associadas a formas base da LC. Para tal, os sistemas com esta abordagem utilizam gramáticas separadas para a análise da LP, para a síntese da LC e para a transformação das estruturas da LP em formas equivalentes na LC (*ibidem*: 503-504).

Apesar de teoricamente não haver um limite estabelecido para o número de regras gramaticais que é possível adicionar ao sistema, na prática existe um limite a partir do qual a introdução de complexidade adicional não se traduz em ganhos de qualidade em termos de resultados de tradução. Em vez disso, conflitos internos e regras contraditórias produzem novos erros arbitrários (Stein, 2013: 5).

Sistemas com esta abordagem, a par da tradução por interlíngua, são considerados sistemas de segunda geração. Tanto esta abordagem como a tradução por interlíngua foram as abordagens mais comuns utilizadas em TA durante as décadas de 70 e 80 (Hutchins, 2003b: 504).

2.2.1.3. Tradução por interlíngua

A abordagem à tradução por interlíngua assume que é possível converter um TP em representações sintático-semânticas comuns a mais do que uma língua (mas não

necessariamente “universais”). Seria a partir destas representações abstratas que o TC seria construído (*ibidem*: 503). Em comparação com os sistemas que utilizam uma abordagem de tradução por transferência, os sistemas que utilizam interlíngua não necessitam da fase de transferência entre duas línguas uma vez que utilizam uma metalíngua para veicular o sentido em ambas as línguas de forma não ambígua. Esta língua universal (interlíngua) seria a LC em qualquer tradução, numa primeira fase, e, numa segunda fase, seria a LP para a construção do sentido numa LC (Stein, 2013: 8). Uma interlíngua pode ser baseada numa língua “lógica” artificial, numa língua auxiliar “natural” (como o esperanto), num conjunto de primitivos semânticos comuns a todas as línguas ou num vocabulário supostamente “universal”. (Hutchins, 2003b: 503)

A tradução por interlíngua é composta por duas etapas: 1) do TP para a interlíngua e 2) da interlíngua para o TC. Na análise do TP são utilizados procedimentos específicos para essa LP particular (não são direcionados para uma LC específica). Da mesma forma, os procedimentos utilizados para a geração ou síntese do TC são específicos da LC e não necessitam de informação da LP (*ibidem*).

Se, por um lado, sistemas que utilizam interlíngua permitem maior economia de esforço por permitirem que, tanto um programa de análise de uma LP específica como um programa de geração para uma LC, possam ser usados mais do que uma vez, por outro, a utilização de uma interlíngua num ambiente multilíngue implica um adensamento da complexidade da própria interlíngua (*ibidem*).

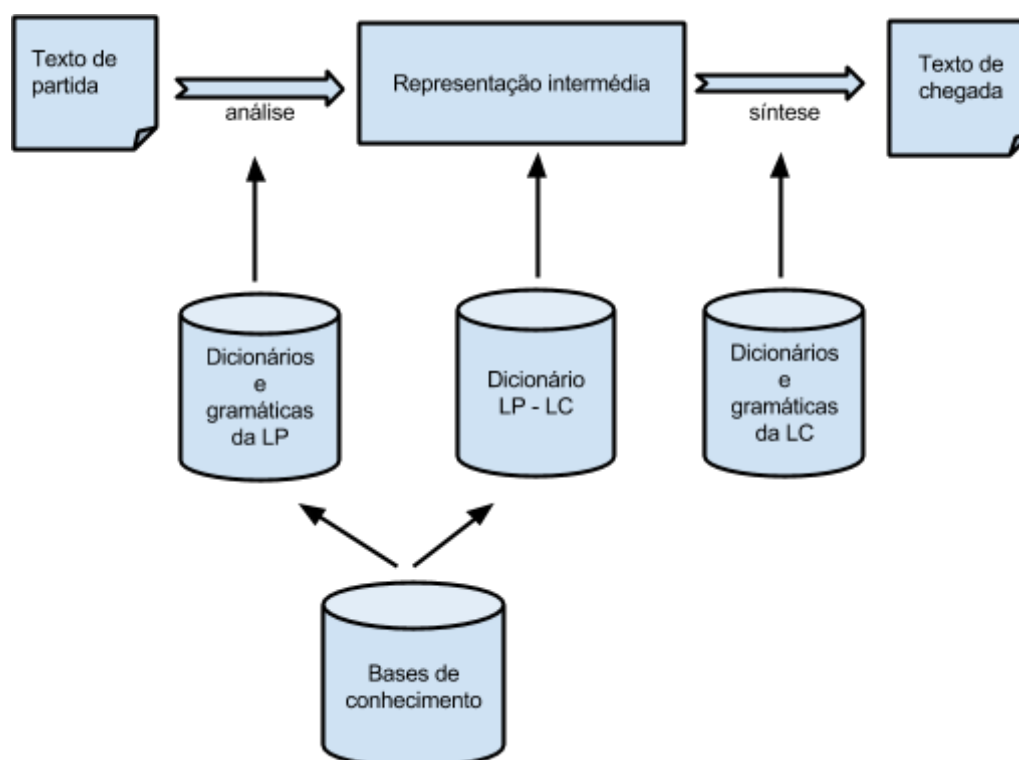


Figura 4 – Representação do processo de tradução por interlíngua (Hutchins, 2007: 18).

Embora tendo sido temporariamente abandonada durante a década de 80 devido à sua complexidade, esta abordagem conheceu renovado interesse no início do século XXI, especialmente com o desenvolvimento de bases de conhecimento compostas por ontologias (redes de conceitos organizados de forma hierárquica) (Hutchins, 2003b: 520).

2.2.2. Tradução baseada em *corpora*

A década de 90 do século passado marcou a emergência de um novo paradigma em termos de abordagens à tradução nos sistemas de TA. A denominada tradução baseada em *corpora* (ou também abordagem empírica³⁰) desenvolveu-se com o aumento exponencial de textos disponíveis na forma de *corpora* paralelos, ou seja, conjuntos de textos com as respetivas traduções. Com a disponibilização de uma grande quantidade de dados de forma organizada, abriram-se novas formas de abordar os problemas relacionados com a TA e que diferiam consideravelmente das abordagens tradicionais baseadas em regras linguísticas. Foi no âmbito desta nova corrente de investigação em TA que surgiram as abordagens à tradução baseada em exemplos e estatística.

³⁰ Cf. Hutchins, 2006: 380

2.2.2.1. Tradução baseada em exemplos

De acordo com Hutchins (2006: 380), o modelo de tradução baseado em exemplos (sigla inglesa EBMT de *Example Based Machine Translation*) foi inicialmente proposto no início da década de 80 no Japão por Makoto Nagao, embora somente tenha sido verdadeiramente desenvolvido uma década mais tarde. Este modelo tem por base a reutilização de traduções existentes para uma nova tradução. Apesar da sua aparente semelhança com os sistemas de memórias de tradução ou CAT (do inglês *Computer Assisted Translation*)³¹, a diferença reside no facto de ser o processo automático a determinar a recombinação dos fragmentos de tradução para dar origem à tradução e não um tradutor humano, conforme acontece em sistemas CAT.

Somers (2003: 514-515) explica que nos sistemas EBMT o processo de tradução é composto por três etapas: correspondência (*matching*), alinhamento e recombinação. Na primeira etapa o sistema procura exemplos capazes de contribuir para a tradução, com base no grau de semelhança com o *input*. Isto pode ser implementado de várias formas: a) os exemplos podem ser guardados em estruturas de árvore anotadas e com ligações entre os constituintes da LP e LC; b) exemplos e *input* são tratados como cadeias de caracteres em que a associação é feita através da comparação entre as sequências; c) os exemplos são anotados com informação morfossintática; d) vários exemplos semelhantes são combinados num exemplo mais geral que contenha variáveis; ou e) a utilização de todas estas abordagens em que o conjunto de exemplos é composto por casos muito gerais que são como regras, casos parcialmente gerais ou exemplos literais. Para evitar exemplos contraditórios ou sobreposição de exemplos, estes são eliminados ou alterados – tal como os sistemas de regras, o sistema EBMT necessita de intervenção humana e, como tal, encontra-se sujeito a inconsistências. Na segunda etapa (alinhamento), são identificadas as partes da tradução que irão ser reutilizadas. Por fim, as partes identificadas são colocadas na devida ordem (recombinação).

³¹ Ambos analisam o TP em segmentos que são comparados com um conjunto de segmentos anteriormente traduzidos de forma a identificar o segmento na LC com o maior grau de semelhança.

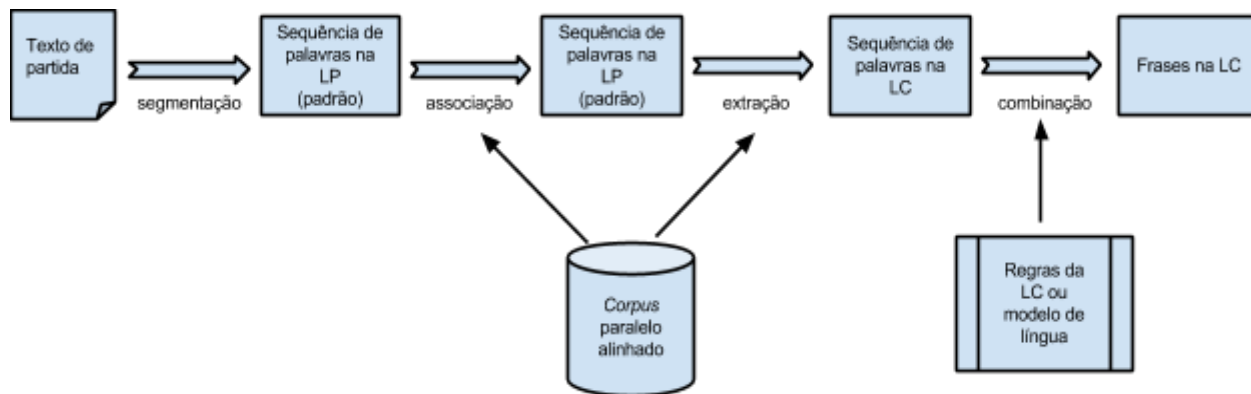


Figura 5 – Representação do processo de tradução baseada em exemplos (Hutchins, 2011: 10).

2.2.2.2. Tradução automática estatística

Na sua forma mais pura, a tradução automática estatística (TAE) não utiliza qualquer informação linguística. O modelo da TAE baseia-se no alinhamento de palavras ou *phrases*³² num *corpus* paralelo e no cálculo da probabilidade de uma palavra numa LP corresponder a uma ou mais palavras numa LC. Este modelo tem na sua conceção a modelização do processo de tradução em termos de probabilidades estatísticas. Desta forma, considera-se que para um par de enunciados f e e ³³ (enunciado de partida e de chegada, respetivamente) há uma probabilidade $P(e|f)$, isto é, a probabilidade de e ser a tradução de f . O processo de tradução resume-se, então, à descoberta do melhor valor para $P(e|f)$ (Somers, 2003: 516).

Somers (*ibidem*: 516-517) refere que o cálculo desta probabilidade irá depender de duas outras medidas. A primeira é a probabilidade de as palavras numa LP corresponderem a outras numa LC (modelo de tradução). A segunda é a probabilidade de essas palavras se combinarem corretamente numa LC (modelo de língua). Ou seja, enquanto o modelo de tradução permite identificar um conjunto de palavras consideradas as mais prováveis para a tradução, o modelo de língua permite estabelecer a organização mais provável desse conjunto de palavras numa LC. Contudo, nem sempre existe uma correspondência

³² *Phrase*, no âmbito da TAE, é entendida como uma sequência de palavras consistentes com um alinhamento lexical, que pode ou não coincidir com um constituinte linguístico (Espana, 2011: 110).

³³ As letras f e e são frequentemente utilizadas na literatura sobre TAE para designar língua de partida e língua de chegada, respetivamente. Como a base textual de apoio ao projeto Candide - que esteve na origem da investigação e dos sistemas TAE atuais - consistia em documentos parlamentares canadianos em inglês e francês, o artigo de Brown *et al.* (1990), considerado seminal nesta área da TA, adotou a nomenclatura e (de *English*) e f (de *French*) para relatar as suas experiências na tradução deste par de línguas. A partir daí vários autores, de entre os quais Koehn (2010), utilizam a mesma nomenclatura para designar língua de partida e língua de chegada. Como tal, também neste trabalho se optou por manter esta nomenclatura.

1:1 entre palavras de duas línguas. Uma palavra em f pode dar origem a uma ou mais palavras em e – problema que foi designado de problema de fertilidade. Além disso, facilmente se percebe também que calcular a probabilidade de todas as combinações prováveis de todas as palavras possíveis numa tradução não é uma tarefa simples, mas antes bastante complexa. Contudo, tal pode ser simplificado calculando as probabilidades de sequências de palavras relativamente pequenas, designadas n-gramas, em que “n” representa um qualquer número de palavras. Ainda assim, mesmo com esta simplificação, é necessário proceder a cálculos complexos para determinar a correta combinação das palavras numa tradução. De forma a simplificar ainda mais este processo, parte-se do princípio de que a ordem das palavras no TC será igual à do TP. Contudo, como se sabe que na transição de uma LP para uma LC existe sempre uma distorção (em maior ou menor grau) na organização das palavras, é igualmente calculado um valor de distorção entre um par de línguas de forma a facilitar o cálculo das probabilidades em termos de organização dos constituintes das frases numa LC. Desta forma, os sistemas de TA estatística, se sabem que uma sequência de palavras, como “a mesa”, ocorrem dessa forma, não é porque reconheça o género das palavras, a sua categoria ou relação gramatical, mas apenas porque reconhecem esta sequência como mais frequente (*ibidem*: 517).

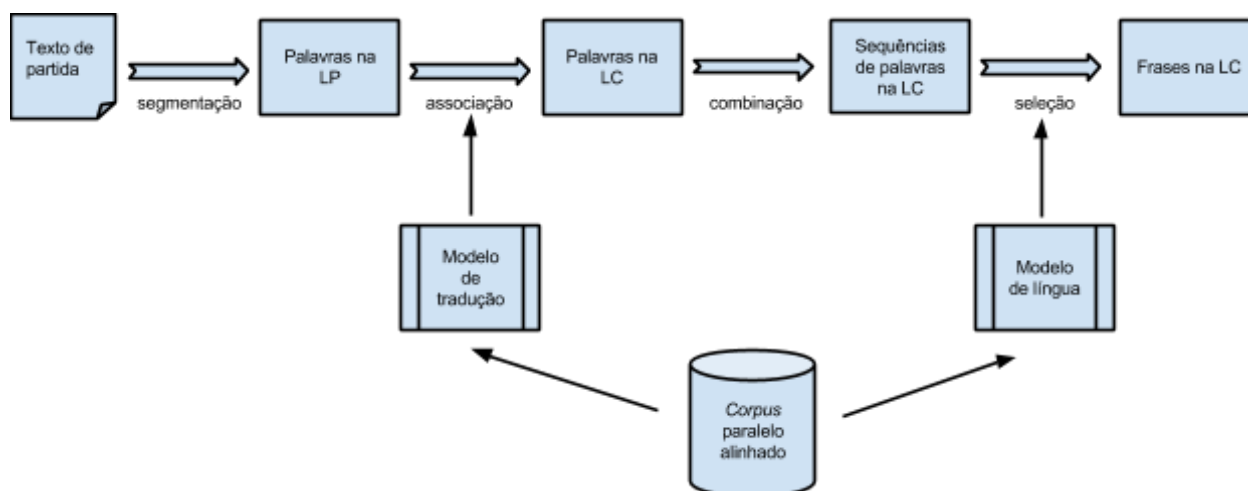


Figura 6 – Representação do processo de tradução estatística (Hutchins, 2011: 9).

Apesar de os primeiros sistemas deste modelo terem apresentado resultados muito promissores, quando comparados com os resultados dos sistemas de regras, a

capacidade de aumento da qualidade da tradução revela-se limitada (*ibidem*). Muitos erros seriam facilmente corrigidos com o mínimo de conhecimento linguístico.

2.2.3. Sistemas de tradução híbridos

Os sistemas híbridos nasceram do reconhecimento dos diferentes pontos fortes e fracos dos sistemas baseados em *corpora* e em regras e da potencialidade da união destes dois tipos de sistemas. Os sistemas híbridos utilizam vários tipos de combinações. Somers (2003: 518) refere que alguns utilizam a componente EBMT para resolver problemas que sistemas de regras não conseguem resolver, outros combinam a fase de análise dos sistemas de regras com a fase de geração EBMT, outros ainda utilizam uma abordagem estatística para a análise do TP e uma abordagem baseada em regras para a fase da transferência e geração do TC. Há ainda sistemas híbridos em que o TP passa por vários processadores de TA que utilizam diferentes técnicas. Ainda de acordo com Somers (*ibidem*), estes sistemas multiprocessadores incorporam um módulo de avaliação para analisar o grau de fiabilidade do resultado da tradução. Por fim, um módulo “moderador” é utilizado para comparar os resultados dos vários processadores, seleccionar aqueles com a pontuação mais alta ou confirmar traduções semelhantes propostas por diferentes processadores ou seleccionar as melhores partes da tradução de cada processador, juntando tudo num só resultado final.

2.3. Avaliação da tradução automática

O que faz com que um sistema ou uma abordagem de tradução seja melhor do que outra(a)? De que forma se orienta a investigação na área de TA para seleccionar o caminho a seguir de entre tantas possibilidades teóricas? É através da avaliação dos resultados empíricos dos sistemas de TA que se distinguem os sistemas melhores e os caminhos mais produtivos na investigação.

A avaliação dos resultados produzidos por sistemas de TA tem sido uma questão prevaiente desde o início da investigação nessa área. A complexidade desta tarefa tem muito a ver com a complexidade da própria avaliação da tradução em si, seja ela de um tradutor humano ou de um sistema de TA. Esta complexidade tem a ver não só com a variabilidade da língua (ou seja, o facto de não existir uma única tradução correta, mas

várias possibilidades), mas também com a própria subjetividade da avaliação realizada por humanos³⁴.

De que forma se pode então avaliar a qualidade de uma TA? Koehn (*ibidem*) refere que esta tarefa tanto pode ser realizada por avaliadores humanos como por sistemas informáticos que comparam o resultado da TA com traduções produzidas por humanos, designadas traduções de referência, por se considerarem estas as de qualidade mais elevada.

O que se espera, então, de uma métrica de avaliação? Koehn (2010: 220-221 e 240) revela que se espera que esta seja: económica (tanto em termos de tempo como de dinheiro); consistente (ou seja, que diferentes avaliadores cheguem às mesmas conclusões com a mesma métrica e que a avaliação de uma parte do *corpus* seja consistente com a avaliação de outra parte); correta (ou seja, que indique o desempenho real dos sistemas e que avalie com melhores resultados os melhores sistemas). Além disso, deve ainda servir para otimizar o sistema de TA e produzir métricas significativas (ou seja, capazes de transmitir um significado claramente interpretável) (*ibidem*).

Além das métricas relativas à qualidade, é também importante avaliar outros parâmetros que distingam os sistemas de TA entre si. Nesse âmbito, é de igual importância a avaliação da rapidez e dimensão do sistema; da facilidade de integração com outras aplicações informáticas; e da capacidade de adaptação a domínios específicos (*ibidem*: 240).

Tudo isto acaba por permitir distinguir os sistemas entre si e orientar os passos a tomar na investigação em TA.

2.3.1. Avaliação humana

Koehn (*ibidem*: 218) refere que uma forma óbvia de avaliar sistemas de TA é utilizar um avaliador humano para produzir juízos acerca dos resultados da TA. Para isso são necessários avaliadores fluentes no par de línguas da tradução ou, pelo menos, fluentes na língua de chegada. Koehn (*ibidem*) refere que este tipo de avaliação é normalmente realizada frase a frase, embora possa ser necessário conhecer o contexto, em casos de documentos mais longos.

³⁴ Koehn (2010: 217) demonstra que os tradutores raramente estão de acordo em relação a uma tradução.

Contudo, acrescenta Koehn (*ibidem*), a forma de avaliar o grau de correção de uma TA é algo muito lato. O mais comum é utilizar o critério da fluência e da adequação. A fluência inclui tanto critérios de gramaticalidade como de seleção lexical. A adequação avalia se o sentido da frase de partida é mantido na frase de chegada. Uma vez que juízos de fluência e adequação estão intimamente relacionados, em vez de classificar estes numa escala absoluta, é normalmente utilizado um sistema de classificação comparativa para avaliar dois ou mais sistemas (*ibidem*: 218-220). Assim, torna-se mais fácil a um avaliador humano avaliar de forma mais consistente se o resultado do sistema A é melhor do que o resultado do sistema B do que classificar cada resultado em termos absolutos em relação à fluência e adequação.

2.3.2. Avaliação automática

Apesar das vantagens da avaliação humana, Koehn (*ibidem*: 222) esclarece que o facto de esta ser realizada por especialistas torna-a não só dispendiosa, como também bastante demorada, quando comparada com outras possibilidades automáticas. Como tal, a avaliação da qualidade dos sistemas de TA tem sido feita principalmente de forma automática, uma vez que, no fundo, o que os investigadores pretendem é saber de forma o mais fiável e rápida possível se uma determinada alteração resultou numa melhoria do sistema. A avaliação automática da TA tem sido um campo em franco crescimento, ao ponto de os investigadores orientarem o seu trabalho de acordo com os resultados deste tipo de avaliação.

De acordo com Koehn (*ibidem*: 222-223), todas as métricas de avaliação automática da TA utilizam um critério comum: a comparação dos resultados da TA com traduções humanas, consideradas de referência. A ideia que se encontra na base deste parâmetro de avaliação é a de que uma tradução que se aproxime mais da tradução de referência tem mais probabilidade de ser considerada correta do que uma que se afasta dessa. Como tal, o desafio é construir uma métrica que avalie o grau de semelhança. A precisão (*precision*) é uma métrica que avalia quantas palavras produzidas pelo sistema TA correspondem à tradução de referência, enquanto a revocação (*recall*) avalia quantas palavras deveriam ser geradas pelo sistema TA e quais destas estão corretas tendo em conta a tradução de referência. Ambas as métricas são importantes na avaliação automática de TA. Não queremos, por um lado, produzir palavras erradas (precisão) e, por outro, não queremos esquecer-nos de nenhuma (revocação). Além disso, queremos

também saber quantos passos – adições, eliminações, substituições – seriam necessários para a produção do sistema de TA igualar a tradução de referência – isto é medido através da denominada taxa de erro por palavra (WER – *Word Error Rate*). Estes são alguns dos métodos comuns a muitos modelos de avaliação automática.

Uma das métricas mais utilizadas é a BLEU (*ibidem*: 226). Tal como outras, esta métrica utiliza uma tradução humana de referência como contraponto. A BLEU calcula a percentagem de n-gramas idênticos entre ambas as traduções (pode utilizar várias traduções de referência de modo a tomar em conta a variabilidade da tradução). Trata-se de uma métrica orientada à precisão, que estabelece uma penalização de brevidade para evitar que traduções demasiado curtas (que possam omitir vários n-gramas de referência) tenham pontuação demasiado alta.

A métrica NIST, também esta uma das mais métricas mais comuns, apresenta-se como uma variação da BLEU, no sentido de procurar suprimir algumas das suas lacunas. Esta métrica, no geral atribui maior valor a n-gramas com mais informação (os mais raros no *corpus*) (Doherty, 2013: 25).

A métrica METEOR é outra das métricas automáticas mais conhecidas e utilizadas. Esta métrica inclui tanto componentes de precisão como de revocação, assim como uma fragmentação de pontuação para considerar a ordem das palavras na frase (*ibidem*: 26). Além disso, considera as semelhanças ao nível do lema e de sinónimos (Koehn, 2010: 241). Devido ao facto de se apoiar em recursos específicos da língua, o METEOR apenas pode ser aplicado a determinadas línguas (Doherty, 2013: 256).

Muitas outras métricas automáticas existem atualmente³⁵ e novas métricas vão surgindo, o que revela que a utilização de métricas automáticas na avaliação dos resultados dos sistemas de TA tem sido objeto de constante debate por parte da comunidade científica. Ainda assim, estudos revelam existir uma correlação elevada entre avaliação humana e automática, ou seja, sistemas avaliados com pontuação baixa por sistemas automáticos, receberam pontuação igualmente baixa por parte de avaliadores humanos (Koehn, 2010: 229).³⁶ Como tal se conclui que, apesar de imperfeitos e com grande espaço para melhorias, as métricas automáticas são extremamente úteis para fornecer informação

³⁵ Vide, por exemplo, listagem disponibilizada em Euromatrix, 2007.

³⁶ Koehn (2010: 228-229) refere que as métricas de avaliação automática são fonte de intenso debate pela comunidade científica. Na verdade, de acordo com este autor (*ibidem*), estes métodos são utilizados quase em exclusivo na investigação da TAE, uma vez que a investigação em sistemas de TA de outras abordagens aponta várias lacunas a este tipo de avaliação. O principal argumento de defesa das métricas de avaliação automática consiste exatamente no facto de se provar que existe uma forte correlação entre resultados da avaliação humana e da avaliação automática (*ibidem*).

relativamente fidedigna e de forma rápida acerca da qualidade de uma tradução por forma a orientar os esforços no desenvolvimento da tecnologia de TA.

2.4. Aplicações de sistemas TA

Quando Weaver expôs as suas ideias acerca da exequibilidade da tradução automática no célebre memorando de 1949, fê-lo tendo em mente uma utilização muito concreta para esta nova tecnologia:

“Even if it would translate only scientific material (where semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worthwhile”

Weaver considerava que a TA poderia ser útil desde logo para ajudar a traduzir (*i.e.* compreender) textos (principalmente de carácter científico) mais fácil e rapidamente, tendo em vista um “futuro pacífico para o planeta”³⁷ (Weaver, 1949: 4). Contudo, desde então as utilidades para a TA expandiram-se e ganharam contornos inimagináveis para a altura.

Hutchins (2010: 1) considera existirem três tipos principais de necessidades de tradução: tradução de elevada qualidade para publicação (assimilação) e tradução de menor qualidade utilizada para transmitir o sentido geral de um texto numa determinada língua de partida (disseminação). Mais recentemente, surgiu uma nova necessidade para a tradução de mensagens de correio eletrónico, salas de conversa virtuais, entre outras formas de comunicação social via Internet, em que a qualidade não é essencial para a compreensão da mensagem que se pretende transmitir (comunicação). Segundo Hutchins (2009: 13), os primeiros sistemas de TA, instalados em supercomputadores, eram utilizados por grandes empresas, entidades militares ou governamentais com o objetivo de produzir traduções que eram depois pós-editadas por tradutores ou revisores humanos. Apesar das inúmeras deficiências no resultado final da TA, já na época os resultados rápidos e consistentes disponibilizados por estes sistemas comprovaram que a TA em larga escala poderia ser rentável em termos de custos. Para melhorar os resultados destes sistemas, no final dos anos 70, foram introduzidos

³⁷ Tradução nossa.

métodos de controlo da linguagem e desenvolvimento de sistemas TA para sublínguas específicas (como, por exemplo, o sistema Méteo para a tradução da previsão meteorológica no Canadá).

Com o desenvolvimento da informática, tanto ao nível do *hardware* como do *software*, e com a massificação dos computadores a partir principalmente da década de 90, a TA começou a ser utilizada também na localização de produtos informáticos. Com a necessidade de colocar no mercado *software* e respetiva documentação de apoio (como manuais de utilizador) em várias línguas em quase simultâneo, a TA surge como a solução óbvia para esta nova necessidade (*ibidem*: 14).

Hutchins (*ibidem*) refere que, inicialmente desenvolvidos para tradutores profissionais, durante a década de 80, os sistemas de TA para computadores pessoais extravasaram o âmbito da tradução profissional e passaram a ser utilizados por tradutores não profissionais com o intuito de produzirem traduções para assimilação. Os sistemas de TA atualmente disponíveis para tradutores não profissionais são, na sua grande maioria, concebidos para lidar com textos de qualquer domínio do conhecimento (contrariamente com o que acontece com muitos sistemas de TA profissionais utilizados por grandes empresas). Devido à própria natureza destes sistemas não profissionais, pouco se sabe da utilização concreta que lhes é dada, embora Hutchins (*ibidem*) considere que o mais provável é que sejam usados por tradutores ocasionais que apenas queiram extrair uma ideia geral de textos estrangeiros e que não deem muita importância à qualidade da tradução. Atualmente, os sistemas de TA são utilizados em paralelo ou incorporados nas próprias ferramentas automáticas de apoio à tradução para auxiliar tradutores profissionais na produção de traduções de elevada qualidade para publicação.

A partir de meados da década de 90, com o sucesso comercial de aparelhos informáticos móveis (como dicionários eletrónicos portáteis, *laptops*, *PDA*, *smartphones*, etc.), cresceu também a oferta de programas informáticos de TA para este tipo de tecnologia. O *software* de TA passa a estar incorporado nos próprios aparelhos e, mais recentemente, os aparelhos móveis passaram a ser usados como terminais de acesso a serviços de TA em linha. Embora os primeiros serviços de TA disponibilizados através da Internet não fossem gratuitos (cf. Hutchins, 2009: 15), a expectativa do aumento de vendas de *software* de TA para PC e de serviços de valor acrescentado de pós-edição levou a que algumas empresas de TA (a Altavista, com o Babelfish, baseado no sistema Systran, foi das primeiras) disponibilizassem serviços de TA em linha com resultados imediatos. Com a entrada em cena da Google, com o serviço em linha Google Tradutor,

que se apresentava com um paradigma completamente diferente (tradução baseada em estatística³⁸, quando até aí os sistemas de TA em linha eram maioritariamente baseados em regras), e que disponibilizava não só um leque muito mais vasto de línguas, mas ainda interação com outras ferramentas em linha (nomeadamente com as páginas em linha, através da tradução instantânea destas), o público em geral passa a ter uma relação muito mais próxima e direta com este tipo de sistemas. Segundo Hutchins (2009: 15-16), pouco se sabe sobre os utilizadores (quem são) e a utilização (quais os objetivos) dos serviços de TA em linha. De acordo com Hutchins (2009: 16) o que se sabe é que muitos donos e desenvolvedores de páginas estão a remeter os utilizadores para serviços de TA em linha para a tradução das suas páginas na Internet (o que demonstra desconhecimento acerca das limitações destes sistemas), embora atualmente já existam produtos comerciais de TA para a localização de páginas na Internet. Além disso, os serviços de TA em linha gratuitos parecem estar também a ser utilizados como dicionários, na medida em que a maioria das pesquisas feitas é de uma ou duas palavras. Contudo, o uso destes serviços tem crescido exponencialmente.

Ainda no domínio da Web, tem-se assistido, embora com sucesso e interesse comercial pouco conhecido, à pesquisa e desenvolvimento (principalmente na Coreia e Japão) de sistemas de TA para mensagens de correio eletrónico (*ibidem*).

Mais desafiante e ainda pouco explorada tem sido a TA de mensagens nas redes sociais. As tentativas realizadas nesta área têm servido para demonstrar as semelhanças deste tipo de textos com a linguagem falada e muitos dos problemas partilhados entre estas duas áreas permanecem ainda por resolver (*ibidem*).

De acordo com Hutchins (*ibidem*: 16-17), a investigação em TA tem trabalhado a par com a investigação na área da tradução da fala. A tradução da fala representa uma mudança radical em relação à TA tradicional (*i.e.* de texto escrito). Contrariamente à linguagem escrita, a linguagem falada é coloquial, elíptica, dependente de contexto, interpessoal e apresenta-se maioritariamente em forma de diálogo. A TA tradicional, por seu turno, tem trabalhado principalmente em linguagem técnica, que é organizada, não ambígua e unipessoal. Alguns dos problemas identificados na tradução da fala podem ser controlados ao restringir a comunicação a domínios específicos (comunicação empresarial, reservas hoteleiras, negociação de datas de reuniões, etc.). Existem já alguns sistemas operacionais na área da medicina (para mediação da consulta

³⁸ Apenas em 2007 a Google passou a usar um sistema baseado em estatística, pois até então utilizava um sistema baseado em regras linguísticas (*vide* subcapítulo 4.2.1. “Google Tradutor”, p. 67)

entre médico e paciente), militar (para comunicações militares) e do turismo, por exemplo. Contudo, são ainda sistemas pouco flexíveis e dedicados a domínios muito específicos.

A investigação tem-se também voltado para a aplicação da TA na legendagem, especialmente com o intuito de servir grupos emigrantes ou línguas minoritárias³⁹. Apesar das dificuldades próprias deste tipo de tradução, a TA enfrenta ainda a escassez de recursos (como, por exemplo, dicionários, *corpora*, corretores ortográficos, etc.) que permitam “alimentar” estes sistemas.

Além de minorias e imigrantes, há outros grupos desfavorecidos da sociedade que estão a ser auxiliados pela TA, de acordo com Hutchins (*ibidem*: 18-19). Recentemente a investigação em TA tem procurado desenvolver sistemas capazes de traduzir língua gestual. A investigação nesta área é ainda muito incipiente e os problemas detetados nesta área vão para além dos problemas encontrados na tradução de textos escritos (combinação complexa de gestos manuais e corporais e expressões faciais).

Ainda de acordo com Hutchins (*ibidem*: 19), outras utilidades que têm sido dadas à TA têm sido a recolha, análise e extração de informação, incorporação de sistemas de pergunta e resposta, criação de resumos e escrita técnica. O acesso a informação multilingue foi, desde sempre, uma das aplicações-alvo da TA, mas à medida que a recolha de informação foi seguindo modelos estatísticos e os sistemas de TA foram seguindo modelos baseados em regras, a relação entre ambos foi diminuindo. Contudo com o desenvolvimento recente dos sistemas de TA estatísticos, a relação entre ambos tem-se estreitado novamente. O mesmo aconteceu com a extração de informação (*text mining*). Entidades públicas e privadas procuram analisar documentos e textos em línguas estrangeiras para extrair informação importante para as suas atividades (comerciais, económicas, espionagem, etc.).

A área da pergunta e resposta é uma área que tem estado ativa dentro do campo da inteligência artificial há vários anos (*ibidem*: 20). O objetivo dos sistemas de TA é recolher respostas em texto numa base de dados em resposta a questões colocadas (preferencialmente) em linguagem natural. A possibilidade de produzir perguntas e respostas em várias línguas tem atraído atenções recentemente.

³⁹ De acordo com Hutchins (2009: 18), a designação de “língua minoritária” pode ser determinada geograficamente (por exemplo, hindu, punjabi e bengali são línguas minoritárias em Inglaterra, onde a língua maioritária é o inglês) ou comercialmente (i.e. línguas que não têm relevância económica ou comercial).

Finalmente, o ímpeto de grandes empresas criarem documentação em várias línguas num curto espaço de tempo tem ajudado a integrar processos de escrita técnica com a tradução. Os sistemas de TA são utilizados para escrever mais rapidamente um rascunho de um texto numa língua que não é totalmente dominada pelo autor (*ibidem*). As aplicações da TA não se esgotam nestas possibilidades, nem os sistemas de TA são utilizados para simplesmente tradução “pura”. Na maior parte das vezes, a TA é incorporada noutros sistemas em apoio à comunicação bilingue ou multilingue, numa gama cada vez maior de situações e num cada vez maior número de sistemas (*ibidem*).

3. Sistemas de Tradução Automática Estatística

A tradução automática estatística (TAE) é uma abordagem que utiliza métodos estatísticos e exemplos de tradução para traduzir textos automaticamente. De acordo com Specia (2013: 74), é o paradigma de investigação em TA dominante na atualidade. A área tem avançado muito desde os modelos tradicionais baseados na palavra, passando pelos modelos baseados em *phrases* e outros modelos mais sofisticados que têm em consideração a estrutura frásica. Atualmente, a tendência recente consiste na incorporação de informação linguística acerca do TP e TC (*ibidem*).

A TAE aborda a tradução como um problema de aprendizagem automática, em que são utilizados algoritmos de aprendizagem num *corpus* paralelo⁴⁰ para o sistema traduzir frases desconhecidas em relação ao *corpus* de treino. O grande interesse, tanto académico como comercial, neste tipo de abordagem, em detrimento de outros, resulta da facilidade de construção de sistemas TAE. Com uma ferramenta TAE e um *corpus* paralelo suficientemente extenso, é possível construir um sistema de TA para um qualquer par de línguas num período curto de tempo de semanas ou mesmo dias (cf. Lopez, 2008: 2). A qualidade destes sistemas irá resultar da quantidade, qualidade e domínio do *corpus* paralelo no qual se baseia (*ibidem*). O grande interesse que hoje existe nos sistemas TAE reside, na opinião de Lopez (*ibidem*: 2-3), essencialmente em cinco fatores:

1. o crescimento da Internet permitiu o acesso a grandes quantidades de texto em diversas línguas e em diferentes domínios;
2. a necessidade de assimilar informação disponível na Internet muitas das vezes numa língua que o leitor não domina;
3. os avanços tecnológicos (tanto em termos de redução de custos como de aumento das capacidades informáticas) permitiu a expansão de sistemas baseados em dados;
4. o desenvolvimento de métricas de tradução automáticas permitiu o desenvolvimento rápido e iterativo de sistemas de TA em geral, e dos sistemas de TAE em particular;

⁴⁰ Um *corpus* paralelo consiste num conjunto de textos e respetivas traduções devidamente alinhados.

5. o desenvolvimento e disponibilização livre e gratuita de ferramentas de TAE, das quais o Moses será a mais conhecida, permitiram o acesso a este tipo de ferramentas para investigação e fins comerciais.

Neste capítulo iremos abordar os conceitos e modelos básicos que compõem os sistemas de tradução automática com base estatística. Explicar, ainda que sucintamente, a constituição e a composição de um sistema de TAE implica, necessariamente entrar no domínio da estatística e da computação. Contudo, sendo este um trabalho realizado no âmbito de mestrado em Linguística, apenas serão apresentados os componentes básicos destes sistemas e os modelos teóricos que estão na sua base, deixando de parte, sempre que possível, algoritmos e fórmulas matemáticas. Isto resulta não só das nossas próprias limitações em termos de conhecimentos em estatística e programação, mas também do âmbito de uma tese de mestrado em linguística. De salientar que, apesar de alguns contributos de outros autores e obras, este capítulo irá basear-se e seguir essencialmente a organização e o exposto por Koehn na obra *Statistical Machine Translation* (2010), por se tratar de uma obra de referência na área da TAE.

3.1. A Teoria do Canal Ruidoso

Na base da abordagem estatística à TA está a aplicação da Teoria do Canal Ruidoso, desenvolvida por Shannon em 1948⁴¹, no âmbito de uma teoria matemática da comunicação. Apesar de ter sido inicialmente relacionada com a TA por Weaver no seu influente memorando de 1949 (cf. Weaver, 1949: 10), somente na década de 90, e após o sucesso na aplicação em sistemas de reconhecimento da fala na década de 70, esta teoria foi aplicada ao domínio da TA (Specia 2013: 74-75). De acordo com a teoria de Shannon, quando um emissor comunica uma mensagem está a codificá-la numa determinada língua (*ibidem*).

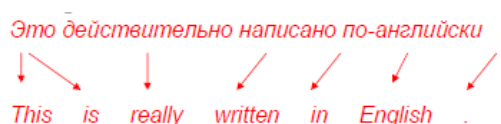


Figura 7 – Exemplo da aplicação da teoria de Shannon à tradução (Specia, 2010: 7).

⁴¹ Shannon, 1948.

Nesse âmbito, ao tradutor cabe a tarefa de decodificar a mensagem numa língua f para posterior reconstituição da mesma o mais fielmente possível noutra língua e .



Figura 8 - Teoria do canal ruidoso de Shannon com base em Specia (2010: 7).

Em termos de probabilidades, traduzir será encontrar o equivalente de uma frase e numa outra língua, ou seja, conforme explicado por Specia (2013: 75), será a procura pelo argumento que maximiza (*argmax*) a probabilidade (P) de recuperar o texto original (e) dado o texto com ruído (f). Isto poderia ser representado da seguinte forma:

$$\operatorname{argmax}_e P(e | f)$$

Aplicando o teorema de Bayes⁴², este problema pode ser decomposto em partes menores e mais simples (*ibidem*). Assim:

$$\operatorname{argmax}_e P(\mathbf{e} | \mathbf{f}) = \operatorname{argmax}_e \frac{P(\mathbf{f} | \mathbf{e})P(\mathbf{e})}{P(\mathbf{f})}$$

Uma vez que f (*i.e.*, o texto de partida) é sempre o mesmo entre todas as possíveis frases e , $P(f)$ pode ser ignorado, resultando em (*ibidem*):

$$\operatorname{argmax}_e P(\mathbf{e} | \mathbf{f}) = \operatorname{argmax}_e P(\mathbf{f} | \mathbf{e})P(\mathbf{e})$$

Conforme referido por Specia (2013: 75), o processo de decomposição do problema de tradução em problemas menores e do modelamento de cada etapa individualmente é motivada pelo facto de ser possível recolher contagens e dados estatísticos mais fiáveis para problemas mais pequenos. Os modelos resultantes da decomposição de $P(e | f)$ correspondem a dois componentes fundamentais de qualquer sistema de TAE: o modelo

⁴² O teorema de Bayes é um teorema da teoria das probabilidades que permite avaliar de que forma a probabilidade de uma teoria ser verdadeira é afetada por novas evidências (*in* <http://www.trinity.edu/cbrown/bayesWeb/>).

de tradução $P(f | e)$ e o modelo de língua $P(e)$. Estes modelos procuram dar resposta àquilo que Jurafsky e Martin (2000 *apud* Ramanathan s.d.: 13) consideram ser a melhor tradução: aquela que é fiel à frase de partida e fluente na língua de chegada. O modelo de tradução $P(f | e)$ procura obter a fidelidade ao original, estabelecendo correspondência lexical entre as línguas de partida e de chegada e utilizando, para isso, *corpora* paralelos (España, 2011 28). O modelo de língua $P(e)$ procura obter a fluência na língua de chegada, utilizando, para o efeito, *corpora* monolíngues na língua de chegada (*ibidem*). Um outro componente, o “descodificador” constrói possíveis traduções de acordo com dados fornecidos pelo modelo de tradução e/ou de língua e seleciona a tradução mais provável (*ibidem*:132).

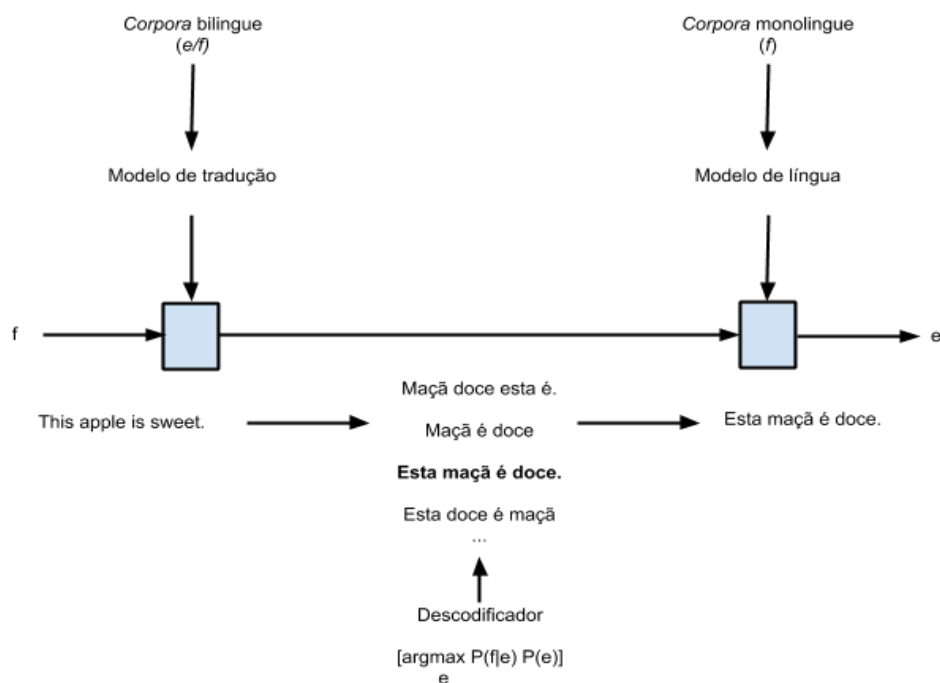


Figura 9 – Esquema da interação entre os componentes de um sistema de TAE com base em Knight e Koehn (2003: 4-5).

Analisemos, de seguida, em maior pormenor os três componentes principais dos sistemas TAE: modelo de tradução, modelo de língua e descodificador.

3.2. Modelo de tradução

Os primeiros modelos de TAE são relativos à tradução ao nível da palavra e resultam da investigação da IBM no projeto Candide, no final dos anos 90 (Koehn, 2010: 81). Embora nos últimos vinte anos a investigação em TAE tenha feito consideráveis progressos, o projeto da IBM lançou as bases daquilo que são os sistemas TAE atuais. Olhemos, como tal, para os primeiros sistemas baseados em palavras e passaremos depois para os sistemas mais avançados, baseados em *phrases*.

3.2.1 Modelos baseados na palavra

Uma palavra pode ter vários significados não só numa única língua, mas também em línguas distintas. “Sweet”, por exemplo, em português pode significar, entre outros, doce, fresco, agradável, perfumado ou melodioso⁴³. Todos estes significados são bastante diferentes e variarão conforme o contexto onde ocorrerem. Koehn (2010: 82) questiona de que forma poderia a estatística ajudar na seleção do melhor termo para traduzir a palavra “sweet”. Se tivermos um conjunto de textos em inglês emparelhados com a respetiva tradução em português (*corpus* paralelo) poderíamos analisar quantas vezes a palavra “sweet” ocorre e verificar não só quais as palavras utilizadas em português para a traduzir, como também recolher o número de ocorrências de cada uma. Assim, poderíamos verificar, por exemplo, que a palavra “sweet” ocorre 1.000 vezes no *corpus* e que foi traduzida 500 vezes como “doce”, 150 como “agradável” e por aí diante. Ou seja, o *corpus* poderia indicar-nos não só quais as possíveis traduções de “sweet” como ainda nos permitiria identificar com que frequência ocorreria cada dessas traduções. De acordo com Koehn (*ibidem*), com estas contagens, seria, portanto, possível fazer uma distribuição de probabilidades de tradução lexicais. Isto, por seu lado, permitir-nos-ia avaliar qual a tradução mais provável em português para a palavra inglesa “sweet” (tendo em conta, naturalmente, apenas o *corpus* em questão e ignorando por completo o contexto onde estas palavras estão inseridas) (*ibidem*).

⁴³ Consulta no dicionário inglês-português da Porto Editora através do sítio da Infopédia em <http://www.infopedia.pt/ingles-portugues/sweet> acedido em abril de 2014.

Formalmente, isto poderia ser representado da seguinte maneira (*ibidem*): $P_f : e \rightarrow P_f(e)$. Esta fórmula representa a probabilidade P de e (palavra na língua de chegada) ser a tradução de f (palavra na língua de partida). No caso de “doce”, sendo que a palavra portuguesa ocorre 500 vezes num total de 1.000 ocorrências de traduções da palavra “sweet”, $P = 0,5$. A este tipo de estimativa, partindo de um conjunto de dados e tendo por base um modelo estatístico, é dado o nome de estimativa por máxima verosimilhança (*maximum likelihood estimation*) por maximizar a verosimilhança ou probabilidade dos dados (*ibidem*: 83).

Implícita na tradução baseada em *corpora* está a noção de alinhamento, ou seja, a associação entre um elemento numa língua f e o elemento correspondente na língua e (*ibidem*: 84).

Por exemplo:

This	apple	is	sweet.
Esta	maçã	é	doce.

Apesar de, neste caso, as palavras em ambas as línguas corresponderem não só em número como em posição, Koehn (*ibidem*: 84-85) lembra que frequentemente as palavras assumem ordenações diferentes em línguas distintas. Isto significa que muitas vezes as palavras têm de ser reordenadas durante a tradução.

Is	this	apple	sweet?
			doce?

Além disso, as línguas podem ainda divergir na quantidade de palavras necessárias para exprimir o mesmo conceito (*ibidem*: 85).

1	2	3	4
The	man	fell	asleep.
O	homem	adormeceu.	
1	2	3	

E há ainda casos em que determinadas palavras não têm um equivalente ou qualquer relação noutra língua e são simplesmente omitidas na tradução (*ibidem*).

The	man	<u>did</u>	not	eat	the	apple.
		/	/	/	/	
O	homem	não	comeu	a	maçã.	

Assim como também há casos em que algumas palavras *e* podem não ter correspondente em *f* (*ibidem*). Nestas situações é introduzido um *token*⁴⁴ nulo que é tratado como outra qualquer palavra *f* (*ibidem*).

[]	Mary	loves	apples.
A	Maria	adora	maçãs.

O *token* nulo é inserido porque se pretende alinhar cada palavra *e* com uma palavra *f*, de forma a ter uma função de alinhamento plenamente definida⁴⁵. Desta forma, é definido um modelo de alinhamento baseado em palavras que permite eliminar, adicionar e duplicar palavras durante a tradução (*ibidem*: 85-86).

Contudo, Koehn (*ibidem*: 86) refere que os textos que constituem um *corpus* paralelo (*i.e.* um conjunto de textos e respetivas traduções) e que servem de base à TAE são alinhados à frase e não à palavra. Além disso, como a maioria das frases apenas ocorre uma vez, mesmo em grandes *corpora*, a modelação da distribuição das probabilidades de tradução não pode ser feita ao nível da frase sob o risco de não se recolher informação útil. Como tal, opta-se por decompor a frase em unidades mais pequenas, neste caso, em palavras. A este processo de decomposição de um processo em pequenas

⁴⁴ *Token* é uma unidade de segmentação utilizada para alinhamento de texto. Pode corresponder a uma palavra, a uma parte da palavra, a um sinal de pontuação, entre outros (*vide* Koehn 34-35).

⁴⁵ Uma palavra *e* é sempre alinhada com outra palavra *f*, incluindo *tokens* nulos. Contudo, uma palavra *f* pode estar ligada a mais do que uma palavra *e* ou a nenhuma (Koehn, 2010: 86).

etapas, modelando essas mesmas etapas com a distribuição de probabilidades, dá-se o nome de modelação generativa.

Num *corpus* paralelo – base para a tradução estatística – temos, então, um alinhamento das frases, mas não das palavras que as compõem e, como tal, não sabemos à partida que ou quais palavras *e* são a tradução de uma determinada palavra *f*, refere Koehn (*ibidem*: 88). Este é um problema típico da aprendizagem automática: o problema dos dados incompletos. Este problema pode ser resolvido através do algoritmo da expectativa máxima (*maximum expectation*) (*ibidem*). Trata-se de “um método de aprendizagem iterativo que preenche lacunas nos dados e treina um modelo através de passos alternados”⁴⁶ (*ibidem*). De acordo com Koehn (*ibidem*), este algoritmo é composto pelas seguintes etapas:

1. inicialização do modelo (normalmente com distribuição uniforme de probabilidades);
2. aplicação do modelo aos dados (fase da expectativa);
3. aprendizagem do modelo a partir dos dados (fase da maximização);
4. iteração das fases 2 e 3 até atingir convergência.

No exemplo seguinte, adaptado de Knight e Koehn (2003: 17) é possível ver como o algoritmo é aplicado ao nível da palavra:

the	house
\	/
a	casa

the	window
\	/
a	janela

a	window
\	/
uma	janela

Inicialmente, desconhecendo a língua *f* (inglês), poder-se-ia dizer que qualquer palavra *e* (português) poderia ser a tradução de uma qualquer palavra *f* (por exemplo, *the*=a ou *the*=janela / *house*=a ou *house*=casa). Como tal, é atribuída a mesma probabilidade a todas as palavras, na etapa da expectativa. No entanto, à medida que o *corpus* vai sendo analisado em maior detalhe (fase da maximização), vai-se verificando que há alinhamentos mais prováveis do que outros: os pares (*the*/a) e (*window*/janela) ocorrem duas vezes, enquanto outros – *a*/uma e *house*/casa – ocorrem apenas uma vez e há ainda pares que nunca ocorrem (é o caso de *house*/janela e *the*/uma). Tendo isto como base, é

⁴⁶ Tradução nossa.

feita nova distribuição das probabilidades. O processo repete-se até haver convergência, ou seja, até haver certeza de que, por exemplo, *window=janela*.

No decurso do trabalho desenvolvido pela IBM no projeto de TAE denominado Candide foram desenvolvidos cinco modelos de complexidade incremental com o objetivo de responder a dificuldades decorrentes do processo de tradução:

1. O Modelo IBM 1 é um modelo generativo para a tradução de frases que se baseia na distribuição de probabilidades de tradução ao nível da palavra (*ibidem*: 86).
2. O Modelo IBM 2 surge para suprir uma lacuna importante do Modelo IBM 1, que se limita a fazer uma distribuição das probabilidades de tradução lexical, deixando de parte a questão da ordem das palavras. O Modelo IBM 2 adiciona um modelo explícito de alinhamento absoluto, ou seja, não só de palavras, mas também da ordem destas na frase. “A tradução de uma palavra f numa posição i numa palavra e em posição j é modelada através de uma distribuição de probabilidades de alinhamento”⁴⁷ (*ibidem*: 97).
3. O Modelo IBM 3 dá conta de grandes transformações num processo de tradução baseado em palavras: tradução de palavras (distribuição de probabilidades de tradução lexical), reordenamento (distorção⁴⁸), inserção de palavras (inserção de *tokens* nulos), eliminação de palavras (palavras com fertilidade igual a 0) e tradução de um-para-vários (fertilidade) (*ibidem*:100-105).
4. O Modelo IBM 4 melhora o modelo anterior através da introdução de um modelo de alinhamento relativo, em que a posição da tradução de uma palavra f é normalmente baseada na posição da tradução da palavra f precedente (*ibidem*: 105-110).
5. O Modelo IBM 5 corrige uma falha dos modelos IBM 3 e 4 ao permitirem que alinhamentos impossíveis tenham probabilidades positivas. Este modelo mantém um registo do número de posição vagas e apenas permite a colocação das traduções nessas posições (*ibidem*: 110-113).

Embora estes modelos IBM não sejam atualmente os modelos de tradução mais avançados, estabeleceram as bases para os sistemas atuais ao introduzirem a noção de

⁴⁷ Tradução nossa.

⁴⁸ A distribuição da probabilidade de distorção difere da distribuição de probabilidades de alinhamento pelo facto de a primeira prever a posição de uma palavra e baseada na posição da palavra f correspondente, enquanto a última prevê a posição de uma palavra f tendo por base a posição da palavra e correspondente (Koehn, 2010: 102).

alinhamento entre palavras (*ibidem*: 81). Contudo, nem sempre é fácil estabelecer um alinhamento entre palavras, chama a atenção Koehn (*ibidem*: 113-115). Um exemplo claro é o das palavras funcionais que nem sempre têm equivalentes noutras línguas. Outro caso é o das expressões idiomáticas, enquanto unidades multipalavras que apenas possuem sentido quando entendidas como um todo. Como tal, nem sempre o alinhamento ao nível da palavra é o mais indicado. Nesse sentido, procurou-se dar uma resposta a essa lacuna utilizando um alinhamento em unidades superiores à palavra: as *phrases*.

3.2.2. Modelos de tradução baseados em *phrases*

Tal como acima mencionado, nem sempre as palavras são a melhor unidade para tradução.

O João leu o livro novo.

John read the new book. ✓

O João leu o livro de novo.

John read the book of new. X

O João leu o livro de novo.

John read the book again. ✓⁴⁹

Atualmente, os sistemas de TAE com melhor desempenho são baseados em modelos de *phrases*, ou seja, unidades formadas por grupos de palavras sem motivação linguística (*ibidem*: 127). Segundo Koehn (*ibidem*: 127-129), estes modelos suprem as lacunas dos modelos de tradução baseados em palavras na medida em que:

- a) as palavras nem sempre são as melhores unidades para tradução, uma vez que uma palavra *f* pode corresponder a mais do que uma palavra *e* e vice-versa;
- b) traduzir grupos de palavras pode ajudar a resolver ambiguidades de tradução;
- c) com um *corpus* grande é possível “aprender” *phrases* úteis cada vez mais longas, até a tradução de frases inteiras;

⁴⁹ Exemplos adaptados de España (2011: 101).

- d) este modelo é conceitualmente muito mais simples, dispensando noções como fertilidade, inserção e eliminação.

Tal como nos modelos IBM orientados à palavra, também as frases são decompostas em unidades menores, neste caso *phrases*. Os pares de *phrases* podem ser extraídos a partir de alinhamento de palavras (e neste caso, os pares de *phrases* selecionados têm de ser consistentes com o alinhamento das palavras) ou diretamente a partir do *corpus*⁵⁰.

Na extração de pares de *phrases* para a criação de uma tabela de tradução de *phrases* é necessário, antes de mais, realizar o alinhamento entre as palavras do par de frases (*ibidem*: 130). Para extrair pares de *phrases* a partir de alinhamentos de palavras é necessário que haja consistência nos alinhamentos das palavras, ou seja, que todas as palavras f na frase F tenham um ponto de alinhamento com palavras e na frase E e vice-versa (*ibidem*: 131). No caso de palavras não-alinhadas, estas tanto podem ocorrer no meio como nas extremidades de uma *phrase* (*ibidem*: 132).

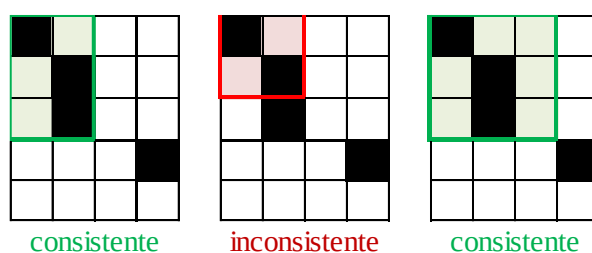


Figura 10 – Todas as palavras devem estar alinhadas umas com as outras. No segundo exemplo, existe uma inconsistência porque há um ponto de alinhamento que se encontra fora do par de *phrases* (Koehn, 2010: 132).

Vejamos o seguinte exemplo⁵¹:

	peter	has	come	back	home	.
o						
Pedro						
regressou						
a						
casa						
.						

⁵⁰ Este último método de aprendizagem de extração de pares de *phrases* não será objeto de análise neste trabalho, uma vez que, tal como referido por Koehn (2010: 148), este tipo de aprendizagem é limitado a pequenos *corpora* e não apresenta melhores resultados do que o método de extração de *phrases* através do alinhamento de palavras.

⁵¹ Exemplo baseado em España (2011: 124).

Dado o alinhamento à palavra entre a frase inglesa (*f*) e a respetiva tradução em português (*e*), seria possível extrair 14 *phrases*: (*Peter*, Pedro) (*Peter*, O Pedro), (*Peter has come back*, O Pedro regressou), (*Peter has come back*, O Pedro regressou a) (*Peter has come back home*, O Pedro regressou a casa), (*Peter has come back home.*, O Pedro regressou a casa.), (*has come back*, regressou), (*has come back*, regressou a) (*has come back home*, regressou a casa), (*has come back home.*, regressou a casa.), (*home*, casa), (*home*, a casa), (*home.*, a casa.), (.,.)

Segundo Koehn (*ibidem*: 131), para cada par de frases, é extraído um conjunto de pares de *phrases*, que tanto podem ser mais longas ou mais curtas, uma vez que tanto umas como outras podem ser úteis. *Phrases* mais curtas têm probabilidade de serem utilizadas com maior frequência, podendo ser aplicáveis a frases nunca antes vistas (como, por exemplo, *home*, casa); *phrases* mais longas, embora menos frequentes, ajudam a captar mais contexto local e, como tal, ajudam a traduzir trechos mais longos de uma só vez (como por exemplo, *has come back*, regressou).

Tal como as palavras numa frase, também as *phrases* sofrem alterações de posição de uma língua para outra. Koehn (*ibidem*: 142-145) refere duas formas de proceder à reordenação de *phrases* na transição de um enunciado de uma LP para uma LC: através do estabelecimento de limites para essa reordenação ou de uma reordenação lexicalizada.

Num modelo em que são estabelecidos limites para a movimentação dos elementos de uma frase na LP para uma LC, a mudança de posição é normalmente punida e apenas é justificada pelo modelo de língua (*ibidem*: 142). Este tipo de modelo de reordenação é utilizado com modelos de língua pequenos (por exemplo, trigramas, ou seja, históricos baseados em três palavras) em que não é possível extrair muita informação para se formarem juízos adequados acerca da gramaticalidade geral de uma frase (*ibidem*). Como tal, nestes casos, ou é utilizada uma tradução monótona (ou seja, em que as posições dos elementos que compõem a frase são mantidas da frase de partida para a frase de chegada) ou é permitida uma reordenação limitada que possibilite, por exemplo, proceder à inversão da ordem sujeito-verbo ou substantivo-adjetivo que ocorre frequentemente entre alguns pares de línguas (*ibidem*: 143). Normalmente, reordenações maiores ou ilimitadas resultam em piores traduções, refere Koehn (*ibidem*).

No modelo de reordenamento lexicalizado, é tida em consideração a orientação do alinhamento de cada palavra que compõe o par de *phrases* (*ibidem*: 143-144). Um ponto

de alinhamento no topo esquerdo significa que a palavra *e* anterior está alinhada com a palavra *f* anterior (orientação monótona). Um ponto de alinhamento no topo direito indica que a palavra *e* anterior está alinhada com a palavra *f* seguinte (orientação *swap*). Se não houver ponto de alinhamento no topo esquerdo ou direito há evidência de uma orientação descontínua.

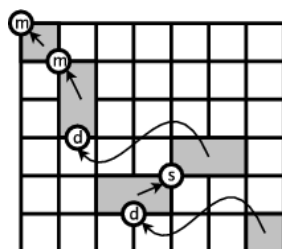


Figura 11 – Tipos de orientação de *phrases* de acordo com o modelo de reordenação lexicalizada: (m) monótona, (s) *swap* (troca) e (d) descontínua (Koehn, 2010: 143).

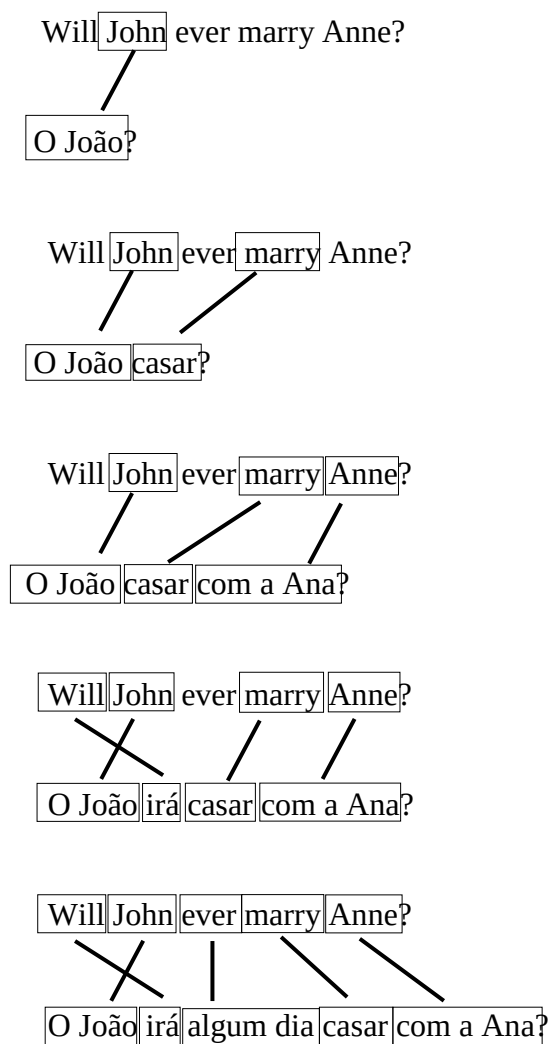
Neste tipo de reordenação, a probabilidade de distribuição de cada tipo de orientação de cada par de *phrases* é estimada a partir da contagem de cada par de *phrases* e de cada tipo de orientação desse par de *phrases* (*ibidem*).

3.3. Descodificador

De acordo com Koehn (2010: 155), no âmbito dos sistemas TAE, a decodificação consiste na tarefa de encontrar a melhor tradução (ou seja, aquela com melhor pontuação) de acordo com o modelo de tradução (seja ele baseado em palavras ou *phrases*). Trata-se de uma tarefa árdua, mesmo para as frases mais simples ou curtas, devido ao número exponencial de opções dada uma frase de partida, uma vez que se trata de avaliar exaustivamente todas as possibilidades, atribuir-lhes probabilidades e selecionar as melhores. Como tal, Koehn (*ibidem*: 155) refere que foram desenvolvidas várias técnicas que permitam encontrar a melhor tradução de forma eficiente – são os denominados métodos de pesquisa heurística. Estes métodos não garantem encontrar a melhor tradução de todas, mas apenas a melhor possível. Há, contudo, dois tipos de erros que podem evitar que a pesquisa heurística consiga encontrar a melhor tradução possível: 1) erros de pesquisa e 2) erros de modelo. Os erros de pesquisa decorrem da incapacidade em percorrer todas as traduções possíveis, enquanto os erros de modelo

resultam em traduções que podem não ser de facto boas, mas que, devido ao modelo, foram classificadas com uma probabilidade elevada Koehn (*ibidem*: 155-156).

Analisando o processo de tradução realizado por um tradutor humano, Koehn (*ibidem*: 156) refere que este tem início com a divisão mental de uma frase em pequenas partes e a seleção da primeira parte a traduzir. Embora uma frase de chegada seja normalmente construída em sequência, não quer dizer que a seleção dos elementos da frase de partida seja feita da mesma maneira. Conforme referido anteriormente, línguas diferentes codificam as posições dos constituintes de uma frase de forma diferente, sendo esse um motivo da reordenação dos elementos da frase aquando da tradução. Como tal, a tradução pode ser feita seleccionando elementos alternados da frase de partida. Atente-se no seguinte exemplo adaptado de Koehn (*ibidem*: 156-157):



Só depois de se terem esgotado os constituintes da frase de partida, é que a tradução pode ser dada por concluída.

A probabilidade de uma tradução é calculada com base na probabilidade do modelo de tradução, do modelo de reordenamento e do modelo de língua (*ibidem*: 158). Contudo, não é apenas no final da tradução que é possível realizar este tipo de cálculo. Durante o processo de tradução é possível calcular a probabilidade da tradução parcial, ou seja, é possível calcular a probabilidade do que já foi traduzido e adicionar as probabilidades das partes da tradução que vão sendo acrescentadas (*ibidem*).

Na tradução de uma frase, o computador é confrontado com várias opções e caberá a este decidir quais utilizar para a tradução (uma palavra ou *phrase* podem corresponder ou ser traduzidos de várias formas) (*ibidem*: 159). Para tal, cria uma lista de possíveis *phrases* e analisa as possibilidades de tradução (*ibidem*). O problema da descodificação consiste em selecionar as opções de tradução corretas e articulá-las numa frase fluente (*ibidem*). De acordo com Koehn (*ibidem*: 159-172) esta tarefa pode ser realizada de diferentes formas:

a) Expansão de hipóteses

Segundo Koehn (*ibidem*: 159-161), tendo em conta a noção de que uma tradução se constrói sequencialmente e através da seleção de opções de tradução, sabemos que iremos construir a tradução da esquerda para a direita, adicionando a tradução de cada *phrase* uma de cada vez. Durante a fase de pesquisa, as traduções parciais construídas são denominadas hipóteses. O ponto de partida é uma hipótese nula, sem qualquer tradução. A partir desta hipótese é construída uma nova hipótese, ou seja, a hipótese é expandida. Isto resultará numa nova hipótese que é relacionada com a hipótese anterior. O processo prossegue desta forma até cobrir toda a frase de partida, ou seja, as hipóteses são expandidas em novas hipóteses, aplicando traduções parciais a traduções de *phrases* *f* que ainda não tinham sido cobertas. Assim que uma hipótese cobre todas as palavras da frase de partida, não pode ser mais expandida. Terminadas assim as hipóteses, procura-se aquela que seja o resultado das melhores traduções parciais com melhores probabilidades.

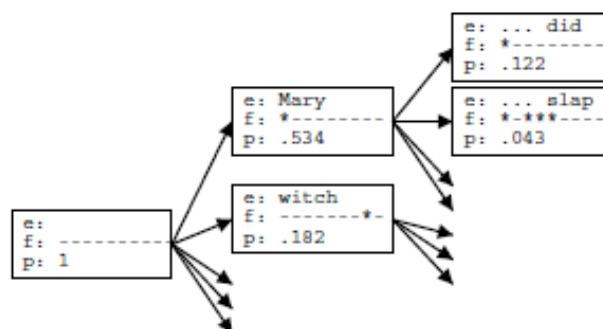


Figura 12 – Diagrama exemplificativo da expansão de hipóteses para a tradução da frase "Mary did not slap the green witch." (Knight e Koehn, 2003: 78)

b) Recombinação de hipóteses

De acordo com Koehn (*ibidem*: 161-162) este meio de decodificação toma em consideração a possibilidade de hipóteses semelhantes seguirem caminhos diferentes. É possível, por exemplo, traduzir duas palavras independentemente ou traduzir as duas palavras diretamente como uma *phrase*.

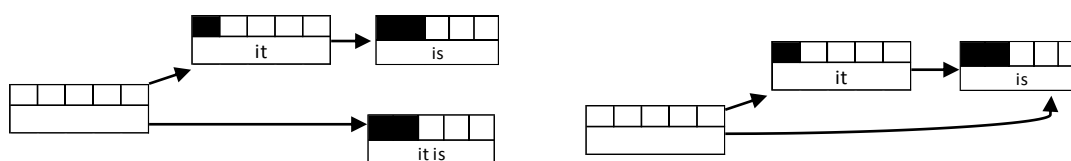


Figura 13 - Exemplos de recombinações de Koehn (2010: 161)

Embora estas duas hipóteses cubram as mesmas palavras f , irão diferir no resultado das suas probabilidades. Neste caso, como uma das duas hipóteses terá pontuação inferior, essa será eliminada, uma vez que se considera que a pior hipótese nunca poderá fazer parte do melhor caminho. Num caminho que contenha a pior hipótese, é possível recombina hipóteses, substituindo a pior hipótese pela melhor, e obter melhores resultados.

c) Decodificação de pilha (*Stack decoding*)

Segundo Koehn (*ibidem*: 163-164), sempre que parecer que no início do percurso existe uma hipótese que não é boa, há a possibilidade de esta ser descartada, ignorando-se assim futuras expansões desta. Contudo, nunca se poderá ter certezas de que essa hipótese não se poderia transfigurar com expansões futuras e dar origem à melhor tradução. Descartar hipóteses demasiado cedo pode levar a erros de pesquisa (apesar de ser possível reduzi-los ao máximo na pesquisa heurística, é impossível eliminá-los por

completo) (*ibidem*: 163). Como tal, para saber se uma hipótese será efetivamente boa ou má, há que levá-la até ao fim. No entanto, tal como referido anteriormente, a investigação sobre todas as possibilidades de tradução é demasiado grande e há, portanto, a necessidade de avaliar hipóteses, compará-las e descartar as piores (*ibidem*: 163).

Koehn (*ibidem*: 164) revela que uma possibilidade é agrupar as hipóteses em grupos (pilhas = *stacks*). Quando estes ficam demasiado grandes, as piores hipóteses são eliminadas. Uma forma de organizar estes conjuntos de hipóteses é de acordo com o número de palavras f traduzidas. Um conjunto inclui todas as hipóteses de tradução de uma palavra f , outro inclui todas as hipóteses de tradução de duas palavras f e por aí diante.

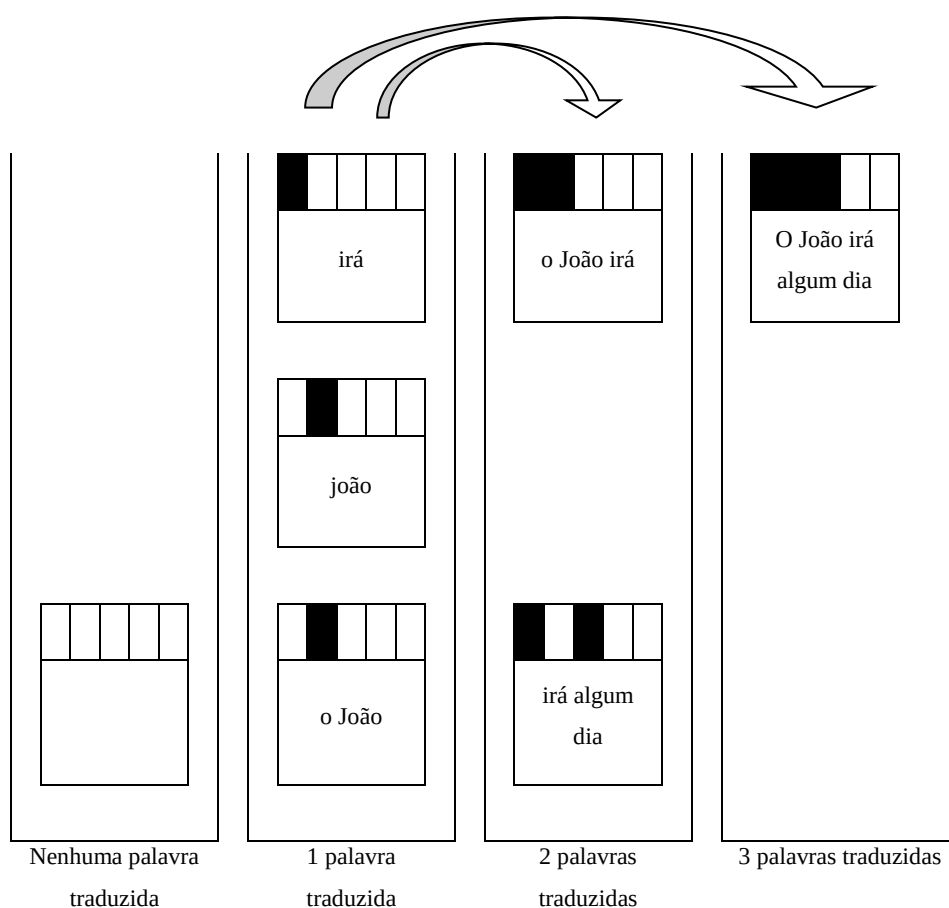


Figura 14 – Hipóteses com o mesmo número de palavras são empilhadas num mesmo grupo. Uma hipótese é expandida ao ser aplicada a outra hipótese, dando origem a uma nova hipótese que é colocada numa nova pilha. (Koehn, 2010: 164)

Esta forma de organizar hipóteses em pilhas e expandi-las a partir destes grupos resulta numa pesquisa heurística bastante compacta (*ibidem*: 165).

Uma vez que cada pilha de hipóteses inclui um determinado número de hipóteses comparáveis (pois cobrem o mesmo número de palavras *f*), é necessário eliminar as piores hipóteses de acordo com a sua probabilidade parcial. Esta ação, designada *pruning* (poda), pode ser realizada de duas formas: através da denominada poda por histograma (*histogram pruning*) ou da poda por limiar (*threshold pruning*) (*ibidem*: 165). Koehn (*ibidem*: 165-166) explica que, enquanto na poda por histograma é definido um número máximo de hipóteses para cada pilha, a poda por limiar explora a noção de que, por vezes, a diferença entre a melhor hipótese a pior hipótese é variável, ou seja, nuns casos esta diferença é grande, noutros nem por isso. Neste caso, pode acontecer que algumas hipóteses próximas da melhor sejam eliminadas na poda por histograma. A poda por limiar propõe estabelecer um limite fixo de diferença entre a melhor e a pior hipótese. Só aquelas hipóteses cuja diferença de probabilidades ultrapasse esse limiar serão descartadas. Este tipo de pesquisa é também, por vezes, denominado pesquisa por feixe (*beam search*), uma vez que este limiar pode ser visto como um feixe com uma determinada amplitude que abrange uma série de hipóteses em cada pilha. Por vezes os sistemas atuais usam ambos os métodos para proceder à eliminação de hipóteses.

Ainda segundo Koehn (*ibidem*: 166-167), uma outra forma de proceder à limitação do crescimento de hipóteses de tradução e, como tal, reduzir a complexidade computacional consiste no estabelecimento de limites de reordenamento. Enquanto alguns pares de línguas não apresentam alterações significativas na ordem dos constituintes da frase no processo de tradução, outros há que requerem alterações de posição significativas. Ao estabelecer limites de reordenamento – ou seja, definir um número máximo de palavras que é possível saltar ao alterar a sequência duma frase durante a tradução – consegue-se um compromisso entre um reordenamento monótono (em que não há lugar a trocas de posição) e um reordenamento desregrado e sem limites. De facto, a investigação tem revelado que além de determinado limite de reordenação (normalmente entre 5 a 8 palavras) os resultados em vez de melhorar, começam a piorar. O estabelecimento de limites permite, portanto, reduzir a complexidade computacional e garantir melhores resultados.

Apesar de existirem outros métodos alternativos de descodificação, estes são alguns dos mais importantes e recorrentes.⁵²

3.4. Modelo de língua

Enquanto o modelo de tradução garante a fidelidade da tradução, ou seja, a correspondência entre uma palavra ou *phrase f* e outra *e*, o modelo de língua serve essencialmente para garantir a fluência da tradução. E fá-lo estabelecendo a probabilidade de uma determinada sequência de palavras ser enunciada numa determinada língua – por outras palavras, avalia a probabilidade de uma frase ser enunciada por um falante dessa língua (Koehn, 2010: 181). No final, pretende-se que um sistema de TAE produza resultados não só fiéis em relação ao texto de partida, mas também fluentes na língua de chegada (*ibidem*).

Na verdade, o modelo de língua num sistema de TAE faz mais do que isso, tem também um papel na decisão relativa à ordem das palavras (*ibidem*).

This house is small.

(a) Esta casa é pequena.

(b) Pequena é esta casa.

(c) É casa esta pequena.*

Embora tanto (a) como (b) sejam frases gramaticalmente corretas em português, o modelo de língua deverá atribuir maior probabilidade a (a) por esta apresentar uma ordem de palavras mais consentânea com a ordem comum das palavras em português.

Além disso, o modelo de língua contribui ainda para a tradução ao selecionar a opção lexical mais provável de acordo com o contexto (*ibidem*).

(1) John is very nice.

(2) That bracelet is very nice.

(1.a) O João é muito simpático.✓

(2.a) Essa pulseira é muito simpática.X

(1.b) O João é muito bonito.X

(2.b) Essa pulseira é muito bonita.✓

⁵² Para mais informação sobre métodos de descodificação alternativos, vide Koehn 2010: 172-176.

Se nalguns contextos, “*nice*” pode ser traduzido como “simpático/a”, noutros deverá ser traduzido como “bonito/a”, por exemplo, sendo, portanto, atribuída melhor pontuação a (1.a) e (2.b).

Koehn (*ibidem*: 182) refere que o principal método utilizado para a modelação da língua é o modelo baseado em n-gramas. Estes modelos baseiam-se em estatísticas acerca da probabilidade de uma palavra seguir outra(s). A palavra “simpática”, por exemplo, certamente que raras vezes seguirá “pulseira é muito”, mesmo num *corpus* muito grande. O objetivo deste modelo é calcular a probabilidade de uma sequência de palavras. Este cálculo é feito mediante a recolha de texto e a contabilização de ocorrências dessa sequência. Contudo, as sequências mais longas de palavras podem nunca ocorrer num *corpus*. Como tal, é necessário decompor esta sequência de palavras em pequenas partes, ou seja, em palavras, para as quais será possível contabilizar as ocorrências e estimar probabilidades de distribuição (*ibidem*). Portanto, “a probabilidade do modelo de língua é o produto das probabilidades das palavras individuais, tendo por base um histórico das palavras precedentes”⁵³ (*ibidem*). Este tipo de modelo em que se avança numa sequência para o estado seguinte (neste caso, palavras) tendo por base um histórico limitado (neste caso, palavras precedentes) é chamado de cadeia de Markov (*ibidem*: 183). O número de estados (palavras) anteriores é denominada a ordem do modelo (*ibidem*). De acordo com a suposição de Markov, uma palavra é afetada por um número limitado de palavras anteriores (*ibidem*). Os modelos com um histórico de três palavras (denominados trigramas), por exemplo, utilizam duas palavras anteriores para prever a terceira palavra. Um modelo de língua pode ser (teoricamente) estimado a partir de um histórico de qualquer dimensão. Para melhor se compreender a ordem do modelo, atente-se no seguinte exemplo de extração de trigramas a partir de um *corpus* constituído apenas por uma frase:

O João gosta de maçãs vermelhas.

(O|Ø, Ø); (João| Ø, O); (gosta|O, João); (de| João, gosta); (maçãs| gosta, de);
(vermelhas| de, maçãs); (.|maçãs, vermelhas)

A estimativa de um trigrama, por exemplo, é realizada contabilizando quantas vezes a sequência p_1 e p_2 é seguida da palavra p_3 , em comparação com outras palavras. O

⁵³ Tradução nossa.

seguinte exemplo, retirado de Koehn (2010: 183), apresenta quais as palavras que seguem mais frequentemente uma sequência de duas palavras no *corpus* Europarl⁵⁴.

the red (total: 225)			the green (total: 1.748)		
Palavra	Contagem	Probabilidade	Palavra	Contagem	Probabilidade
cross	123	0,547	paper	801	0,458
tape	31	0,138	group	640	0,367
army	9	0,040	light	110	0,063
card	7	0,031	party	27	0,015
,	5	0,022	ecu	21	0,012

Contudo, não raras vezes é necessário ter em consideração um histórico (ou seja, um número de palavras precedentes) mais longo (*ibidem*). O número de palavras do histórico é definido de acordo com a dimensão do *corpus* de treino – mais dados implica a possibilidade de utilizar históricos mais longos (*ibidem*).

De um modelo de língua espera-se que dê uma probabilidade alta a sequências de palavras que sejam fluentes na língua de chegada e probabilidade baixa a sequências de palavras pouco fluentes na língua de chegada (*ibidem*: 184). Esta é a noção que está na base da métrica de avaliação comum a modelos de língua, a perplexidade (*ibidem*). Dado um determinado modelo de língua treinado num *corpus* de treino, queremos avaliar a perplexidade de uma frase não incluída no *corpus* (*ibidem*). A perplexidade tem por base a quantidade de massa de probabilidade que é atribuída a palavras observadas (*ibidem*: 185). Um bom modelo de língua não dispensaria probabilidade a sequências impossíveis de palavras numa língua, de forma que houvesse mais massa de probabilidade disponível para sequências possíveis e a maioria estivesse concentrada nas sequências mais prováveis (*ibidem*). Modelos de ordem mais elevada revelam-se melhores a prever palavras, resultando numa perplexidade menor (*ibidem*).

De acordo com Koehn (*ibidem*: 186), a perplexidade é baseada no conceito de entropia. A entropia mede a incerteza numa distribuição de probabilidades. Se uma distribuição tiver resultados com elevado grau de certeza, a entropia é baixa. Quantos mais eventos forem possíveis com probabilidade significativa, maior será a entropia. Como tal, sequências mais longas são normalmente menos prováveis do que sequências mais curtas.

⁵⁴ O Europarl é um *corpus* paralelo extraído das sessões do Parlamento Europeu.

E se um n-grama nunca visto num *corpus* de treino aparecer numa frase teste que se pretende avaliar? N-gramas deste género teriam uma probabilidade zero, esclarece Koehn (*ibidem*: 188). Contudo, isto não será muito útil na prática, porque o facto de uma sequência nunca ter sido vista não quer dizer que não exista ou que esteja errada ou não seja fluente (*ibidem*: 189). Como tal, será útil que estes casos tenham alguma probabilidade (*ibidem*).

À medida que aumenta a ordem do n-grama aumenta também a probabilidade de esse n-grama nunca ter sido observado no *corpus* de treino (*ibidem*). De que forma é, então, possível ajustar as contagens empíricas com as contagens previstas de n-gramas num texto nunca antes observado? Koehn (*ibidem*) sugere que a forma mais simples consiste em adicionar um valor fixo (por exemplo, 1) a cada contagem, eliminando, desta forma, as contagens de 0 – é o denominado método de alisamento adicionar-um (*add-one smoothing*). Contudo, uma vez que o somatório das probabilidades tem de resultar em 1, é necessário ter estas ocorrências imaginárias de n-gramas ao estimar probabilidades. Nesse caso, qual o valor correto a atribuir? Adicionar o valor de 1 dá demasiado peso a n-gramas nunca vistos. Assim sendo, a solução passará por experimentar valores até encontrar aquele que otimiza a perplexidade.

Ao observar um n-grama um determinado número de vezes no *corpus* de treino, o que quer isso dizer? Com que frequência se espera encontrá-lo no futuro? Para responder a essa questão é necessário um segundo *corpus*, acrescenta Koehn (*ibidem*: 190). A solução para arranjar um segundo *corpus* passa por dividir o *corpus* de treino inicial em dois: um para o treino e outro para a validação das contagens. A esta técnica é dado o nome de *deleted estimation*. Contudo, ao dividir a meio um *corpus*, perdem-se dados para contagem, adianta Koehn (*ibidem*: 191). Para resolver este problema, a solução passa por alternar entre o *corpus* de treino e o de validação, *i.e.* um é usado para treino e outro para validação e depois trocam (*ibidem*). O nome *deleted estimation* (estimativa eliminada) provém exatamente do facto de parte do *corpus* ser deixado para validação (*ibidem*: 192).

O objetivo de um ajustamento das contagens consiste em avaliar o significado da ocorrência de um n-grama um certo número de vezes num *corpus* (*ibidem*). Com que frequência se espera encontrar o mesmo n-grama no futuro? Para tal é necessário ajustar ou alisar (*smoothing*) a contagem real à contagem expectável (*ibidem*). Este ajuste ou alisamento é conseguido através da recolha de contagens de estatísticas acerca de quantos tipos de n-gramas ocorreram no *corpus* de treino, quantos n-gramas ocorreram

duas, três ou mais vezes e quantos não ocorreram sequer (*ibidem*). Este método de alisamento é denominado *Good-Turing*.

Tal como já referido, n-gramas de ordem elevada têm em consideração um contexto mais alargado e conduzem a melhores modelos de língua (*ibidem*: 196). Contudo, em *corpus* de treino mais pequenos, estes podem nunca ser observados. Além disso, pode haver casos em que n-gramas de ordem menor, podem ser mais úteis do que n-gramas de ordem maior. Tomemos os seguintes n-gramas retirados de Koehn (*ibidem*) como exemplo: “Scottish beer drinkers” e “Scottish beer eaters”. Koehn (*ibidem*) refere que ambos os n-gramas teriam a mesma pontuação, caso nunca tivessem sido observados num *corpus* de treino, embora o mais fluente seja o primeiro, devido à contradição semântica do segundo. Neste caso, o contexto adicional proporcionado por “Scottish” não acrescenta valor ao n-grama, criando um problema de escassez de dados para avaliar este n-grama. Contudo, os bigramas “beer eaters” e “beer drinkers” poderiam ser, porventura, mais úteis, na medida em que teriam mais probabilidade de aparecerem num *corpus*. Desta forma, Koehn (*ibidem*: 197) alerta para o facto de que, embora n-gramas de ordem mais elevada possam fornecer mais contexto, n-gramas de ordem mais baixa podem ser mais robustos. Portanto, a utilidade de um e de outro tipo de n-gramas pode variar de acordo com a necessidade. Nesse sentido, é útil combinar modelos de língua baseados em n-gramas de maior ordem com modelos baseados n-gramas de ordem menor (*ibidem*). A isto é dado o nome de interpolação. Sempre que um n-grama é observado no *corpus* de treino, é possível calcular probabilidades para prever palavras (*ibidem*: 198). Caso contrário, será aconselhável recuar para n-gramas de ordem mais baixas e com mais dados estatísticos (*ibidem*).

O alisamento de Witten-Bell chama a atenção para a importância da diversidade de palavras previstas. Se, tal como referido por Koehn (2010: 199), nos deparássemos com dois bigramas com histórico composto por “*spite*” e “*constant*”, respetivamente, que ocorressem, cada um deles, 993 vezes, qualquer método de alisamento baseado unicamente na contagem de ocorrências trataria estes históricos da mesma forma. Contudo, observando com maior atenção estes bigramas, poderíamos verificar que apenas quatro palavras diferentes seguem “*spite*” e 415 palavras diferentes seguem “*constant*”. Isto demonstraria que há muito mais incerteza na previsão das palavras que seguem “*constant*” do que “*spite*”. Através deste método de alisamento é reservada maior ou menor massa de probabilidade de acordo com a diversidade das possíveis extensões do histórico (*ibidem*: 199).

Além da diversidade de palavras previstas é igualmente importante considerar a diversidade de históricos. Os modelos de n-gramas de ordem baixa são utilizados sempre que os históricos de modelos de n-gramas de ordem mais elevada são raros ou inconclusivos (*ibidem*: 200). Tomando novamente um exemplo referido em Koehn (*ibidem*), o unigrama “*york*” ocorre, por exemplo, 477 vezes num *corpus*. Contudo, verifica-se que esta palavra ocorre quase sempre precedida de “*new*” – 473 vezes – (formando a palavra “*new york*”, o nome da cidade dos EUA) e raramente de outras palavras – apenas 4 vezes com outras palavras. Isto significa que dificilmente se espera que “*york*” seja usado como segunda palavra num bigrama nunca visto. Por isso, num modelo de retrocesso (também denominado *back-off*) para um unigrama, queremos que a palavra “*york*” tenha muito menos probabilidade do que a sua contagem sugere. Ou seja, queremos que seja tomada em conta a diversidade do histórico. É isso mesmo que o método de alisamento Kneser-Ney faz. De acordo com Koehn (*ibidem*: 200-201), este método, em vez de tomar em conta as 477 ocorrências da palavra “*york*”, utiliza a contagem muito menor de quatro históricos de uma palavra. Se contrastarmos isto com a palavra “*indicates*”, que apresenta as mesmas 477 ocorrências, mas com 172 históricos de uma palavra diferentes, temos uma probabilidade de retrocesso para unigrama muito mais elevada para “*indicates*” do que para “*york*”. Outros modelos mais avançados existem atualmente, contudo, estes constituem a base dos sistemas de TAE.

3.5. A adição de conhecimento linguístico

O processamento puramente estatístico ao nível superficial das palavras e das *phrases*, por vezes não permite dar conta de alguns problemas que seriam mais facilmente resolvidos se houvesse informação mais profunda, ou seja, informação linguística. Dada a evolução de ferramentas capazes de extrair informação linguística a partir de texto⁵⁵, não admira que se tenha considerado complementar as plataformas estatísticas com este tipo de instrumentos ou processos de adição de informação linguística para superar vários problemas a que a abordagem meramente estatística ao nível superficial da palavra e/ou *phrase* não consegue responder. A anotação sintática facilita, por exemplo, a descrição do reordenamento das palavras e *phrases* na frase, o tratamento de palavras

⁵⁵ Tais como analisadores sintáticos, etiquetadores de categorias gramaticais, entre outros.

funcionais (eliminação, adição e tradução), a desambiguação ou a detecção de problemas gramaticais a longas distâncias (Specia, 2010: 101-102). Por outro lado, as ferramentas de anotação sintática apresentam ainda uma taxa de erro bastante elevada e a adição de informação sintática torna os modelos mais complexos, o que aumenta a taxa de erro de pesquisa (*ibidem*: 103)

A inclusão de informação linguística pode ser realizada sem integração real com a plataforma estatística, através da utilização de ferramentas de anotação gramatical antes ou depois do processamento realizado pela plataforma de TAE. Alternativamente, o conhecimento linguístico pode ainda ser integrado no processamento realizado pelo sistema de TAE ao nível de palavra e da *phrase* numa extensão em que a palavra é substituída por um vetor de fatores (Espana, 2011: 177).

A adição de informação linguística num *corpus* antes do processamento do mesmo pelo sistema de TAE pode ser útil, por exemplo, na tradução de línguas ricas morfologicamente. Este tipo de línguas possui uma grande quantidade de vocabulário, o que acarreta muitas vezes problemas relacionados com falta de dados, uma vez que o sistema necessita de detetar as variações (resultantes de flexão ou composição) de uma palavra várias vezes para as poder reconhecer (Specia, 2010: 85). O sistema pode, por exemplo, conhecer “corruptível”, mas este conhecimento não servirá de base para o reconhecimento de “incorruptível”, se esta palavra não constar do *corpus* utilizado no treino, uma vez que o conhecimento que possui é meramente superficial. A solução pode passar por pré-processar o *corpus* de treino para separar as palavras nos seus elementos constituintes (morfemas) e catalogá-los. Estes tanto podem ser ignorados (se não forem realizados na línguas de chegada) ou filtrados (Specia: 2010: 85).

Ela mora numa casa grande

Ela	mor + a	em + (um+a)	casa	grande.
She	live+s	in a	big	house

Neste caso, a palavra “numa” é dividida nos seus constituintes e a flexão no género feminino é eliminada na tradução para o inglês devido à inexistência desta nessa língua de chegada.

Como tal, seria mais simples traduzir a frase se esta fosse pré-processada da seguinte forma:

Ela morar+[3P-SGL] em um grande casa.

O mesmo acontece com palavras compostas: estas podem ser divididas numa fase de pré-processamento para serem depois traduzidas separadamente, conforme é possível verificar no exemplo de Specia (2010: 86):

Aktionsplan = (Akt+ion) + s [elemento de ligação] + Plan > Plano de ação

O pré-processamento do *corpus* com informação linguística pode ainda ser útil na reordenação dos elementos da frase na tradução. Conforme referido anteriormente, as línguas não estruturam os elementos da frase da mesma maneira, nem isto é muitas vezes resolvido de forma adequada por processos meramente estatísticos. De acordo com Specia (2010: 88), uma possibilidade consiste em fazer uma análise sintática do TP, colocar por extenso a estrutura da árvore sintática resultante da análise e aplicar regras de reordenação de forma que a ordem da frase de chegada se aproxime da estrutura habitual da LC. A aprendizagem de regras de ordenação pode ser feita através da análise de um *corpus* paralelo ou baseada apenas na análise sintática (*ibidem*).

Em vez de lidar com os problemas sintáticos antes da fase de processamento da tradução é igualmente possível tratar destes numa fase posterior. Pode-se pegar na melhor tradução candidata e corrigir apenas nesse momento os problemas gramaticais ou pegar num conjunto de melhores resultados, aplicar modelos baseados em informação linguística e reordenar novamente a lista (por exemplo, analisar gramaticalmente os resultados e reordená-los de acordo com a probabilidade de representarem uma boa árvore sintática (Specia, 2010: 90)).

Original: I will give you the apples.

Tradução 1: Eu dou-te a maçã.

Tradução 2: Eu dou-te este maçãs.

Tradução 3: Eu dar-te as maçãs.

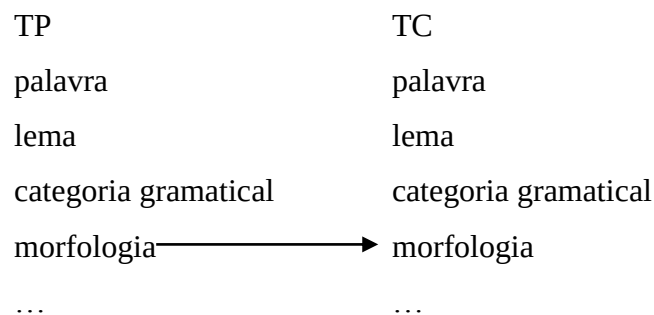
Tradução 4: Eu te as maçãs.

Tradução 5: Eu dou-te as maçãs.

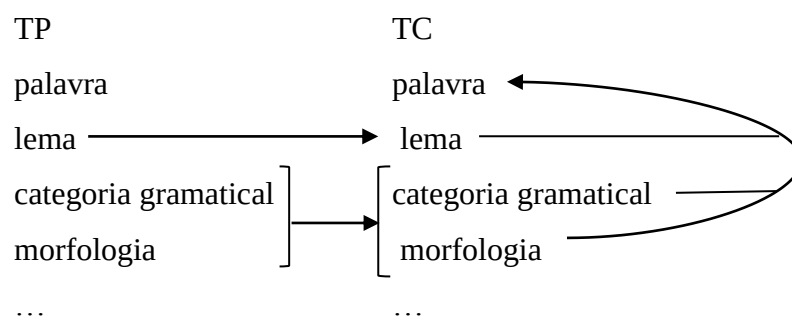
Neste caso, seria possível observar que em termos de coerência de número, a tradução 1 não preserva o plural do original; em termos de concordância na tradução 2 o determinante demonstrativo “este” não concorda em género e número com o

substantivo “maçãs” e na tradução 3 o verbo não concorda com o sujeito; e em termos de sintaxe a tradução 4 teria poucas probabilidades de representar uma boa árvore sintática na LC devido à falta de verbo. Como tal, a tradução 5 seria aquela considerada a mais provável.

Além disso, Specia (2010: 94) refere ainda que é possível integrar conhecimento linguístico durante o processo de treino/pesquisa, adicionando anotações linguísticas ao nível da palavra. Os modelos utilizados nesta fase são denominados modelos fatorizados logarítmico-lineares. Neste tipo de modelo, a palavra deixa de ser apenas um *token* para passar a ser um vetor de fatores que representam informação linguística em diferentes níveis.



Em termos morfológicos, por exemplo, o modelo de língua irá tratar a palavra “boneco” de forma totalmente independente em relação à palavra “bonecas”, apesar de esta última ser a forma feminina e plural da primeira (*ibidem*: 95). Através de um modelo fatorizado logarítmico-linear é possível traduzir o lema (bonec-) e a informação morfológica (-a, feminino e -s, plural) de forma separada e combinar esta informação no resultado final (*ibidem*: 96).



Os modelos fatorizados incorporam informação definida estatisticamente e não apenas linguisticamente motivada (*ibidem*). Estes modelos são decompostos em diferentes

etapas, ora traduzindo os fatores da LP em fatores da LC ora gerando fatores da LC adicionais a partir de outros fatores da LC já existentes (*ibidem*: 97). Este processo pode ser representado da seguinte maneira, por exemplo:

forma superficial *bonecas* | lema *bonec-* | categoria gramatical *NN* | número *plural* | gênero *feminino*

traduzir lema da LP em lema da LC (*bonec-* > *doll, dumm-, puppet*)

traduzir fatores morfológicos e categorias gramaticais (*NN* | *plural-feminino* > *NN* | *plural, NN* | *singular*)

gerar formas superficiais a partir de lema e de fatores linguísticos (*doll* | *NN|singular* > *doll, dummy* | *NN|plural* > *dolls, dummies, ...*)

Dadas as várias opções em cada etapa (o que reflete a ambiguidade da tradução), cada *phrase* de partida pode ser expandida numa lista de opções de tradução (Specia, 2010: 99).

Por tudo, isto é comum que os sistemas TAE atuais adotem estas e outras abordagens baseadas em informação linguística, uma vez que permitem apresentar melhores resultados especialmente na tradução entre línguas distantes (Specia, 2010: 107)

4. Métodos e Materiais

Neste capítulo iremos abordar os métodos e materiais usados para o estudo de caso. Primeiro serão apresentados os sistemas de TAE selecionados para proceder à tradução. De seguida, serão apresentados e descritos os textos que serviram de base a este estudo e, por fim, os instrumentos de avaliação dos resultados de TA.

4.1. Sistemas de TA testados

Para este estudo de caso foram selecionados dois sistemas de TAE: o Moses for Mere Mortals e o Google Tradutor.

A opção pelo Moses for Mere Mortals teve que ver com o facto de esta solução ser de acesso livre⁵⁶ e por se basear no sistema de tradução estatística Moses, porventura, o sistema de TAE livre com melhores resultados e mais explorado por investigadores de todo o mundo. Além disso, o MMM é uma ferramenta que pode ser configurada livremente pelo utilizador, o que permite seleccionar o material com que o sistema será treinado. Apesar de ser uma derivação do Moses, contrariamente a este, não realiza treino fatorizado, ou seja, utiliza apenas modelos estatísticos e não incorpora conhecimento linguístico.

A opção pelo Google Tradutor teve a ver com o facto de ser um sistema de base estatística de livre acesso, fácil de utilizar, que utiliza um conjunto de *corpora* imenso e muito diversificado, uma vez que o objetivo é servir a tradução de qualquer tipo de texto.⁵⁷

⁵⁶ Disponível para descarregamento livre em <https://code.google.com/p/moses-for-mere-mortals/> (último acesso em julho de 2014).

⁵⁷ A Google oferece pouca informação acerca do seu sistema de TA. A informação disponibilizada pela empresa encontra-se aqui: http://translate.google.com/about/intl/en_ALL/. A referência aos *corpora* de treino que suportam o sistema é relativamente genérica – documentos da ONU, livros e páginas na Internet. Contudo, nada é referido acerca de conhecimento linguístico. Contudo, ao possuir um grupo exclusivamente dedicado ao Processamento de Linguagem Natural (<http://research.google.com/pubs/NaturalLanguageProcessing.html>) para prestar apoio aos produtos desenvolvidos pela Google, apesar de não ser divulgado de forma explícita, certamente que o Google Tradutor beneficia de muitas das ferramentas de conhecimento linguístico desenvolvidas por este grupo.

4.1.1. Moses for Mere Mortals

Um dos sistemas de TAE mais conhecidos e explorados atualmente é o sistema de *software* livre Moses⁵⁸. Este sistema, publicado na Internet em 2007, foi desenvolvido no âmbito dos projetos da União Europeia EuroMatrix, TC-STAR (Quadro 6) e EuroMatrixPlus, LetsMT, META-NET e MosesCore (Quadro 7) e com o apoio de várias universidades europeias e norte-americanas e da DARPA (Defense Advances Research Projects Agency) dos EUA⁵⁹.

Visto tratar-se de um *software* livre, ou seja, desenvolvido pela comunidade científica com o objetivo de ser alvo de contributos de investigadores de todo o mundo, a interface com o utilizador é, nas palavras de Machado *et al.* (2010: 23), “um pouco opaca para os não especialistas”. Por esse facto, Machado *et al.* (*ibidem*) justifica a necessidade de simplificar não só o modo de utilização como também de instalação deste sistema, tendo, para o efeito, João Rosas, Hilário Fontes e Maria José Machado desenvolvido uma aplicação “que põe a TA ao alcance de um maior número de utilizadores” (*ibidem*) – o Moses for Mere Mortals (MMM). O MMM, tal como o Moses, é uma aplicação de *software* livre que pode ser descarregada gratuitamente da Internet. Este sistema é treinado no par de línguas desejado através do processamento de “uma quantidade significativa de textos (*corpora*) traduzidos com essa combinação linguística” (*ibidem*: 24). O MMM é composto por um conjunto de *scripts* informáticos que, de acordo com Machado *et al.* (2011: 13), permitem automatizar as tarefas de instalação, criação de um conjunto de ficheiros de demonstração para teste, treino, tradução e avaliação automática. O *script create* permite compilar o sistema Moses e os respetivos pacotes com um único comando; o *make-test-files* permite extrair um pequeno *corpus* de teste a partir do *corpus bilingue* de base; o *train* realiza o treino do sistema a partir do *corpus* de treino; o *translate* realiza a tradução; o *score* faz a avaliação automática de um documento traduzido tendo como base uma tradução de referência; e o *transfer-training-to-another-location* permite transferir os dados do treino para diferentes locais no mesmo computador ou para outro computador (*ibidem*).

⁵⁸ Cf. Machado *et al.*, 2010: 25.

⁵⁹ Fonte: <http://www.statmt.org/moses/>

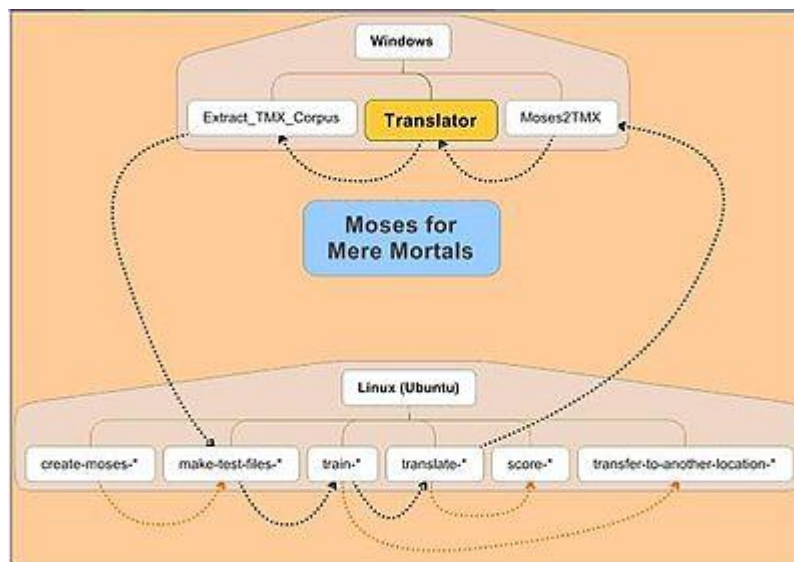


Figura 15 – Diagrama da integração dos diferentes componentes do MMM (Machado *et al.* : 1).

O MMM é instalado e executado em ambiente Linux, no entanto, integra um conjunto de aplicações paralelas de apoio que são executadas em ambiente Windows: Extract_TM_X_Corpus – uma aplicação para a conversão de ficheiros em formato TMX (padrão utilizado pelos sistemas de memórias de tradução) em ficheiros paralelos e alinhados para o “treino” do sistema de TAE; Moses2TMX – uma aplicação para a conversão do ficheiro com o TP e o ficheiro MMM com o TC em ficheiros com a extensão TMX e alinhados para utilização em sistemas de memórias de tradução; “nonbreaking_prefix.pt” – um ficheiro que permite ao sistema ignorar os pontos utilizados nas abreviaturas em português de forma a não confundi-los com ponto final.

4.1.1.1. Instalação

Para a instalação do sistema MMM foi utilizado um computador com 4 GB RAM, 140 GB de espaço de disco e 2 processadores de 2 Ghz. Uma vez que o MMM foi concebido para ser executado numa plataforma Linux, foi instalado o sistema operativo Ubuntu⁶⁰. A versão Ubuntu instalada foi a 10.04 *desktop*, uma vez que a versão MMM a instalar (1.16) apenas é compatível com as versões 10.04, 10.10 e 11.04 Ubuntu. Depois de implementado o sistema Ubuntu, foi instalada a versão 1.16 do MMM, que é possível descarregar gratuitamente da Internet⁶¹. Juntamente com o MMM, foram ainda

60 O Ubuntu é um sistema operativo de código fonte aberto baseado em Linux. A versão instalada encontra-se disponível em <http://old-releases.ubuntu.com/releases/10.04.3/> (último acesso em julho de 2014).

61 Disponível em <https://code.google.com/p/moses-for-mere-mortals/downloads/list> (último acesso em julho de 2014).

descarregadas duas pequenas aplicações que ajudam a fazer a transição entre os ambientes Linux e Windows: o Extract_TMX_Corpus e o Moses2TMX.

Para a instalação do MMM e execução dos diferentes passos que compreendem o processo de criação de um motor de tradução, tradução de documentos e avaliação automática dos resultados foram utilizados os *scripts* que acompanham a versão descarregada da Internet, com a exceção dos *scripts* “create-1.36” e “make_test-tuning_file-0.15” (ambos gentilmente cedidos por Hilário Fontes) que representam a versão atualizada dos *scripts* disponibilizados em linha.

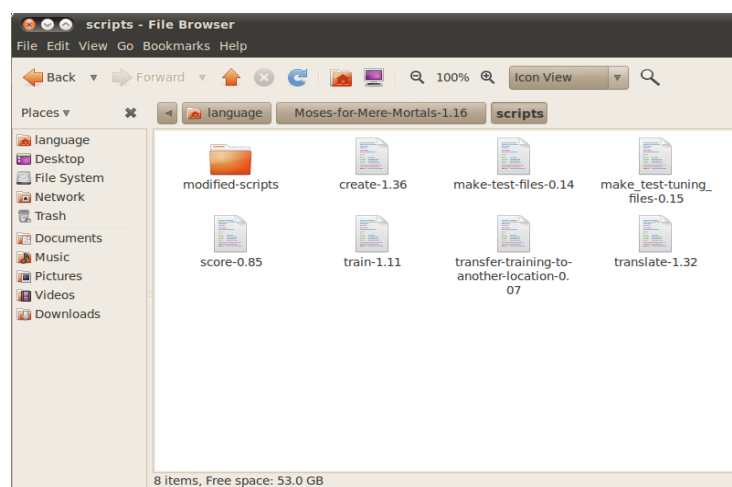


Figura 16 – *Scripts* na pasta Moses-for-Mere-Mortals-1.16.

Todos os *scripts* são divididos em duas áreas:

1. uma primeira área onde se encontram anotações com instruções e informação explicativa acerca de cada um dos parâmetros que podem ser configurados pelo utilizador;

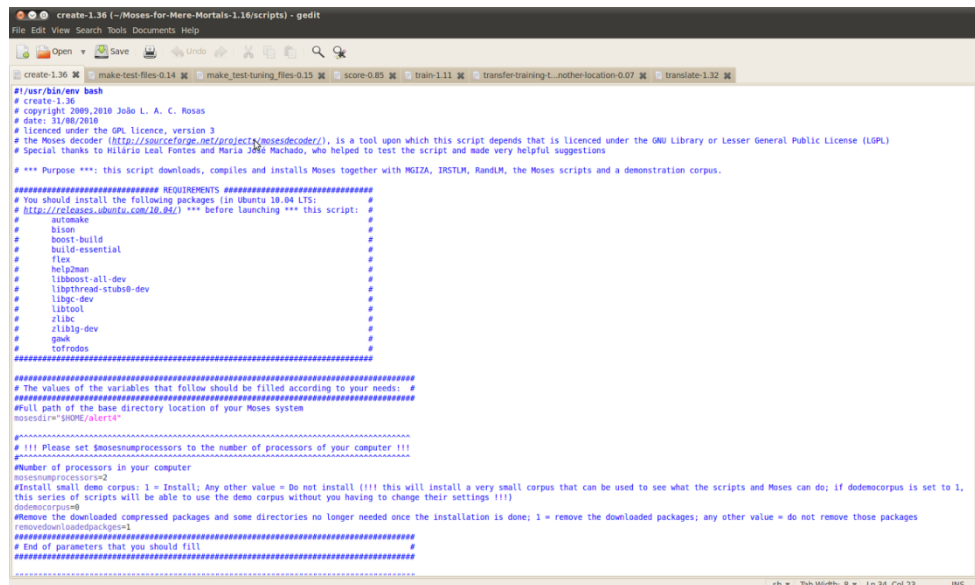


Figura 17 – Área de configuração de parâmetros do script “create-1.36”.

2. e outra, desaconselhada ao utilizador menos avançado, que é constituída pelas linhas de código que irão gerar as ações definidas pela seleção dos parâmetros na área anterior e algumas anotações informativas.

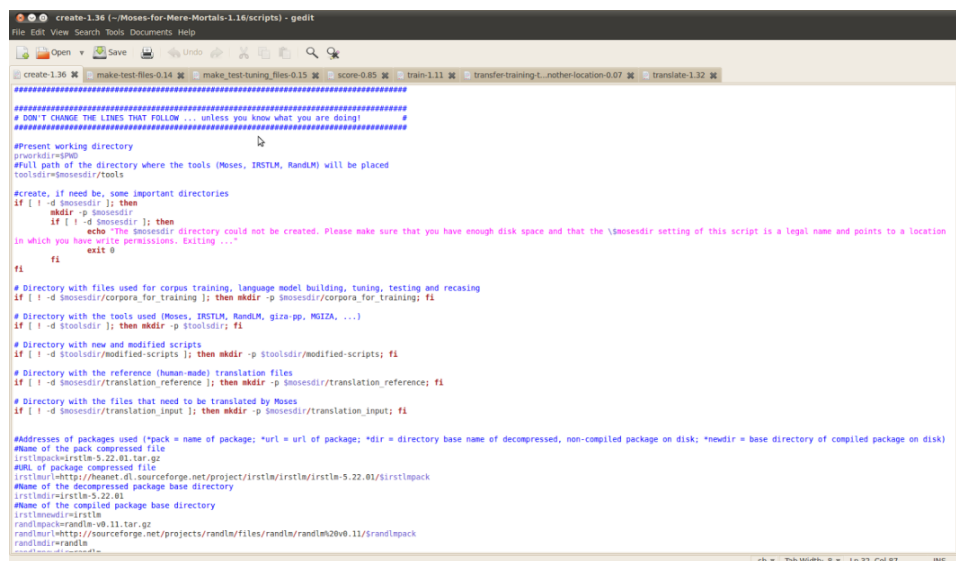


Figura 18 – Área desaconselhada a utilizadores menos avançados e que é constituída pela linha de código que executa as ações definidas pelos parâmetros selecionados pelo utilizador.

Foram criadas 3 instalações MMM – alert, alert2 e alert3 – que foram treinadas com diferentes *corpora* para avaliação de resultados de acordo com os *corpora* utilizados. Em todas as instalações foram utilizadas as mesmas configurações de parâmetros nos

diferentes *scripts* de forma a permitir que os resultados pudessem ser comparáveis. Foram utilizados os seguintes *scripts*:

- “create-1.36” para a criação das instalações MMM;
- “make_test-tuning_files-0.15” para a criação de excertos do *corpus* de treino para treino, teste e *tuning*;
- “train-1.11” para o treino do “motor” de tradução tendo por base o *corpus* bilingue;
- “translate-1.32” para a tradução de um TP para um ficheiro em formato TXT ou TMX;
- “score-0.85” para a avaliação automática BLEU e NIST da tradução tendo por base o TP e uma tradução de referência.

O *script* “make-test-files-0.14” não foi utilizado, por ser a versão desatualizada do “make_test-tuning_files-0.15”. Da mesma forma, o *script* “transfer-training-to-another-location-0.07” também não foi utilizado, mas porque não foi necessário mudar os ficheiros de treino para outra localização no âmbito deste projeto.

Além dos *scripts* foram ainda usadas as aplicações Extract_TMX_Corpus_1.043 (para a extração de ficheiros de *corpus* para MMM com base em ficheiros em formato TMX) e Moses2TMX_1.032 (para a extração de TMX a partir dos ficheiros produzidos pelo MMM).

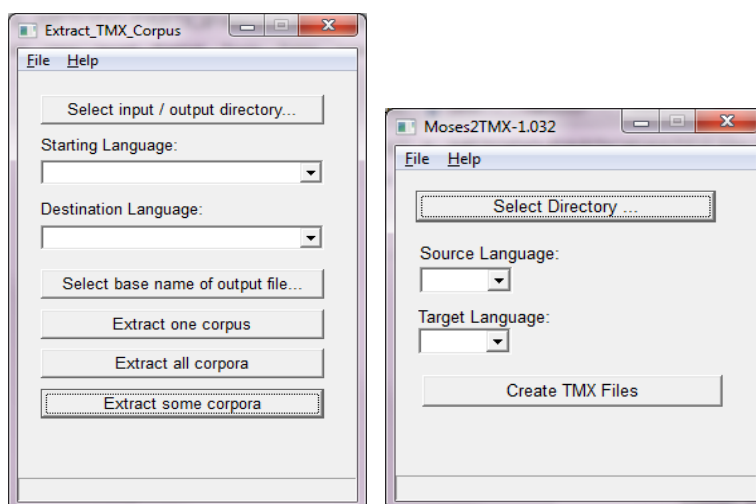


Figura 19 – Aplicações Extract_TMX_Corpus_1.043 (esquerda) e Moses2TMX_1.032 (direita).

4.1.1.2. *Corpora* bilingues de treino para MMM

Sendo o MMM uma ferramenta de TAE que deve ser “alimentada” pelo utilizador, é necessário disponibilizar *corpora* paralelos para o sistema produzir motores de tradução. No âmbito deste projeto, foram utilizados vários *corpora* paralelos provenientes de diferentes fontes. Estes *corpora* encontram-se exclusivamente em português e inglês e foram agrupados da seguinte forma:

- **CorpusALLALL:** este *corpus* é composto por 238.725 segmentos (4.523.233 palavras) alinhados em português de Portugal e inglês dos EUA⁶² e foi extraído a partir das memórias de tradução da ALERT produzidas nos sistemas SDL Trados 2007 e SDL Trados Studio 2014. Estas memórias incluem conteúdo de localização dos produtos clínicos e não-clínicos ALERT, textos de ajuda dos produtos ALERT, conteúdos clínicos incluídos ou traduzidos no âmbito dos produtos ALERT, listas de terminologia clínica, notícias sobre a empresa, textos técnicos e comerciais sobre os produtos ALERT, material didático sobre os produtos e serviços ALERT, textos de âmbito financeiro, legal, gestão de projeto e comercial traduzidos no âmbito das operações correntes da empresa. Apesar de poder conter alguns segmentos em português do Brasil, estes são residuais no conjunto dos segmentos que constituem o *corpus* de português.
- **CorpusALLALLEMEA:** este *corpus* é constituído pelo *corpus* CorpusALLALL e pelo *corpus* da EMEA⁶³ – *European Medicines Agency*, a entidade reguladora da União Europeia no âmbito da avaliação de medicamentos. O *corpus* de treino da EMEA é composto pela documentação emanada por esta instituição, ou seja, o seu âmbito é clínico, embora focado principalmente na área farmacêutica. O *corpus* de treino EMEA foi selecionado para integrar este projeto visto ser o maior e mais fiável *corpus* de treino disponível de forma gratuita no âmbito das ciências da saúde. O *corpus* desta instituição foi descarregado em formato TMX a partir do sítio da OPUS⁶⁴

⁶² O motivo pelo qual foram utilizadas memórias com inglês dos EUA tem a ver com o facto de as memórias com inglês do Reino Unido serem consideravelmente mais pequenas (cerca de 1/10 das memórias com inglês dos EUA) e, como tal, terem material considerado insuficiente para o projeto.

⁶³ Tiedemann, 2009.

⁶⁴ A OPUS é uma página na Internet (<http://opus.lingfil.uu.se/>) que consubstancia o projeto com o mesmo nome, cujo objetivo é disponibilizar de forma livre e gratuita um conjunto de textos traduzidos, convertidos para formatos padrão (tal como o TMX) e alinhados de forma a poderem servir a comunidade de investigadores na área da Linguística Computacional.

(<http://opus.lingfil.uu.se/EMEA.php>) e é composto por 316.070 segmentos (10.540.575 palavras) alinhados em português de Portugal e inglês⁶⁵. Como tal, o CorpusALLALLEMEA é composto no total por 554.795 segmentos alinhados (15.063.808 palavras).

- **CorpusALLALLEMEAEuroparl:** este *corpus* é constituído pelo CorpusALLALLEMEA e pelo *corpus* Europarl⁶⁶. O Europarl é um *corpus* paralelo extraído a partir das sessões plenárias do Parlamento Europeu⁶⁷ disponibilizado através do sítio da OPUS (<http://opus.lingfil.uu.se/Europarl.php>). Este *corpus* é composto por 1.960.407 segmentos (99.959.093 palavras) alinhados e a seleção do mesmo teve por base o facto de se tratar de um *corpus* muito extenso, sem um domínio de conhecimento específico (portanto, de âmbito geral). Tal como no *corpus* EMEA, o Europarl utiliza português de Portugal e inglês do Reino Unido. No total, o CorpusALLALLEMEAEuroparl possui 2.515.202 segmentos (115.022.901 palavras) alinhados.

O facto de os *corpora* bilingues utilizados no MMM conterem tanto variantes do inglês dos EUA como do Reino Unido tem a ver com o facto de:

- a) a tradução realizada na ALERT ocorrer na sua grande maioria (mais de 90%) para e de inglês dos EUA, visto ser a variante adotada oficialmente pela empresa. Como tal, as memórias de tradução de ou para inglês do Reino Unido são demasiado reduzidas para poderem servir de *corpus* para um motor de TAE;
- b) os recursos disponíveis na Internet de forma livre e gratuita são em grande parte fornecidos por instituições europeias, que seguem normalmente a norma do inglês do Reino Unido⁶⁸. Na impossibilidade de arranjar *corpora* bilingues no domínio clínico e da mesma variante de inglês que tivessem dimensão suficiente para serem utilizados em sistemas TAE, optou-se por utilizar os *corpora* acima discriminados, mesmo contendo variantes de inglês distintas. Este fator não foi considerado impeditivo, visto que as diferenças ao nível da terminologia clínica entre os dois tipos de inglês são essencialmente ortográficas, sendo as diferenças

⁶⁵ Não há referência à variante de inglês utilizada, mas tratando-se de uma agência situada no espaço europeu, depreende-se que utilize o inglês do Reino Unido como padrão.

⁶⁶ Tiedermann, 2012.

⁶⁷ Koehn, 2005.

⁶⁸ Vide Directorate-General for Translation, 2014: 1.

terminológicas relativamente poucas (comparando, por exemplo, com o português europeu e o português do Brasil) e identificáveis.⁶⁹

O português presente nos diversos *corpora* é português de Portugal, e a ortografia corresponde ao acordo ortográfico de 1945. O português do *corpus* criado a partir das memórias de tradução é igualmente de Portugal e a ortografia é maioritariamente do acordo de 1945, embora contenha já segmentos com a ortografia do acordo de 1990.

4.1.2. Google Tradutor

O Google Tradutor é uma ferramenta de TAE desenvolvida pela Google, Inc. disponível em 80 idiomas ⁷⁰ e acessível de forma gratuita através de Internet no sítio <https://translate.google.pt/>.

Até outubro de 2007 o Google Tradutor era baseado na ferramenta de TA por regras SYSTRAN⁷¹. A partir dessa data o Google Tradutor passou a utilizar um sistema de TAE desenvolvido pela própria empresa que utiliza uma abordagem estatística.⁷²

Por se tratar de uma ferramenta em linha, não necessita de instalação, embora existam várias aplicações para dispositivos móveis que facilitam o acesso ao Google Tradutor⁷³. A tradução é feita de forma instantânea ao inserir um texto no editor da língua de partida (que pode ser selecionada pelo utilizador ou detetada automaticamente pelo sistema); o resultado da TA é apresentado numa janela paralela da LC (selecionada pelo utilizador).

⁶⁹ Esta constatação tem por base a nossa experiência de tradução na área da Medicina em ambas as variantes do inglês.

⁷⁰ Fonte: http://translate.google.pt/about/intl/pt-PT_ALL/ (último acesso em julho de 2014)

⁷¹ Fonte: http://pt.wikipedia.org/wiki/Google_Translate

⁷² *Ibidem*.

⁷³ Vide <https://play.google.com/store/apps/details?id=com.google.android.apps.translate>.

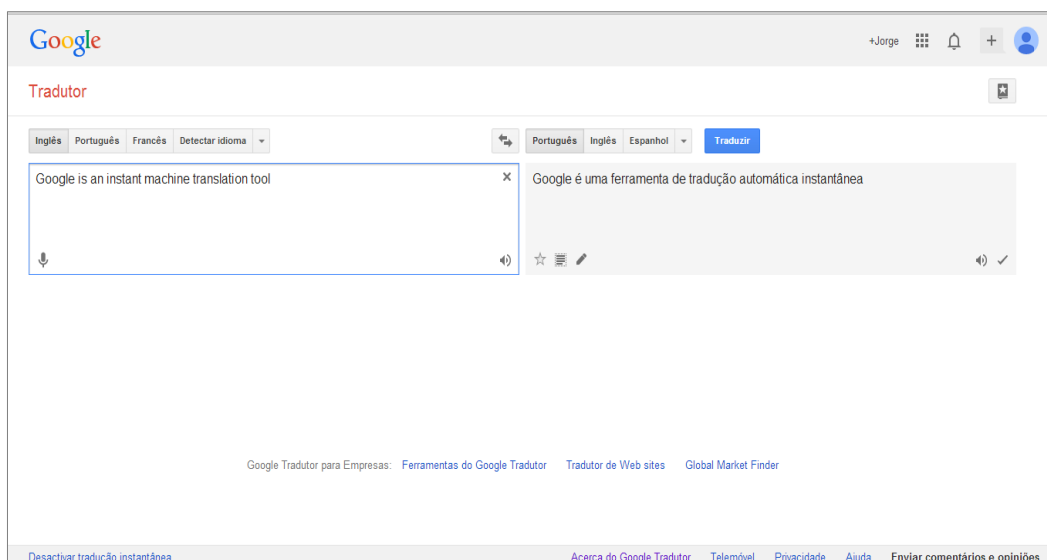


Figura 20 – Página do tradutor automático instantâneo Google Tradutor.

Sendo um sistema de TAE, utiliza vários *corpora* para extrair os modelos de língua e tradução necessários à tradução. Os *corpora* que estão na base do motor de tradução do Google Tradutor não são claramente identificados na informação disponibilizada pela Google⁷⁴, embora outras fontes relatem que o Google utilizou, pelo menos no seu início, *corpora* das Nações Unidas e União Europeia⁷⁵. O sistema é ainda alimentado por memórias de tradução criadas e carregadas na ferramenta de apoio à tradução da Google – o Translator Toolkit⁷⁶ – e melhoradas através de contribuições dos utilizadores tanto do Google Tradutor como do Translator Toolkit. Desta forma, os *corpora* que servem de base ao motor de tradução do Google Tradutor são de carácter geral e não são orientados a um domínio específico do conhecimento.

Contrariamente ao MMM, embora não seja explicitamente declarado pela empresa, é muito provável que o Google Tradutor utilize estratégias de adição de conhecimento linguístico não só através da utilização de Wordnets para vários idiomas⁷⁷, como

⁷⁴ “Quando o Google Tradutor gera uma tradução, procura padrões em centenas de milhões de documentos para determinar qual a melhor tradução.” In http://translate.google.pt/about/intl/pt-PT_ALL/ (último acesso em julho de 2014)

⁷⁵ “Och, who speaks German, English and some Italian, feeds hundreds of millions of words from parallel texts such as Arabic and English into the computer, using United Nations and European Union documents as key sources.” in “Google seeks world of instant translations”, Reuters, 2007 in Wikipedia (http://en.wikipedia.org/wiki/Google_Translate#cite_note-36).

⁷⁶ Translator toolkit - <https://translate.google.com/toolkit/list?hl=en#translations/active> (acesso disponível apenas com conta Google) (último acesso em julho de 2014).

⁷⁷ Vide http://translate.google.pt/about/intl/pt-PT_ALL/ (último acesso em julho de 2014).

também através de vários projetos desenvolvidos pela divisão de Processamento de Linguagem Natural da Google⁷⁸.

4.2. Textos clínicos ALERT®

Os textos selecionados para tradução são textos utilizados na aplicação clínica ALERT®. Estes textos foram criados por profissionais de saúde – médicos, enfermeiros e técnicos – e são disponibilizados na aplicação ALERT® como textos pré-definidos que os utilizadores podem selecionar para facilitar a documentação de informação e evitar a duplicação de dados.

The screenshot displays the ALERT® clinical application interface. At the top, there is a header with the patient's name, 'João Costa SC', and a red 'Urgência' (Emergency) status. Below the header, there is a navigation bar with icons for 'Voltar', 'Alertas', 'Documentação', 'Prescrições', 'Alta', 'PDMS', 'Enfermagem', 'PCE', and 'Gestão de paciente'. The main content area is divided into two columns. The left column contains a list of tabs: 'Resumo', 'Diagnósticos da alta', 'Medicação e outros', 'Pedidos de referência', 'Eventos futuros', 'Instruções ao paciente', 'Informação para outro médico', 'Alta médica', and 'Nota de alta médica'. The 'Instruções ao paciente' tab is selected, and its content is displayed in a red-bordered box. The right column contains a list of predefined text templates, including 'Alergias e problemas', 'Análises', 'Episódios e arquivo', 'Medicação', 'Notas', 'Procedimentos', and 'Sinais vitais e outros'. The bottom of the screen features a navigation bar with icons for 'Retroceder', 'Imprimir', 'Adicionar', 'Pesquisa avançada', 'Cancelar', 'Textos frequentes', 'Importar', 'Atalhos', 'Ajuda', 'Ajuda de contexto', 'Detalhe', and a date/time display showing '18 27 14 Jul 2014' and an 'ok' button.

Figura 21 – Exemplos de textos pré-definidos utilizados pelo médico na atribuição de alta no ALERT®.

⁷⁸ Vide <http://research.google.com/pubs/NaturalLanguageProcessing.html> (último acesso em julho de 2014).

O primeiro documento – “Instruções de alta” (daqui em diante, IA) – consiste num conjunto de textos mais frequentes que o utilizador da aplicação clínica ALERT® (neste caso, apenas o médico ou enfermeiro) pode seleccionar aquando da atribuição da alta a um paciente. Este conjunto de textos contém informação específica acerca do diagnóstico, terapia e acompanhamento ao paciente. Todos os textos possuem um título, normalmente com o nome do diagnóstico atribuído ao paciente, uma explicação acerca da doença, fatores de risco, tratamento ou comportamentos a adotar para a diminuição de fatores de risco e instruções sobre seguimento por parte de profissionais de saúde. Este conjunto de textos foi criado originalmente em inglês dos Estados Unidos, por profissionais de saúde desse país, e validado pela equipa de profissionais clínicos da ALERT, responsável pelos conteúdos clínicos integrados na aplicação clínica ALERT®. Os textos foram disponibilizados em formato Excel e inseridos no SDL Trados Studio 2014 para a segmentação do texto. Depois de segmentado, o texto foi copiado para o editor de texto Notepad++ com a designação “instrucoesdealta.en” para poder ser traduzido para português pelo MMM. A tradução para português de Portugal foi realizada na ferramenta de apoio à tradução SDL Trados Studio 2014. Esta mesma tradução serviu como referência à avaliação automática realizada pelo MMM. Do documento total traduzido, foram selecionados 100 segmentos simultâneos (correspondentes a 1.142 palavras) com menor semelhança com os segmentos que compõem a memória de tradução ALERT, que serviu de base aos *corpora* de treino utilizados pelo MMM. Os segmentos que compõem o IA constituem textos completos, ou seja, possuem sequências com coesão. Estes segmentos são, na sua grande maioria, frases completas, ou seja, com a estrutura mais comum em português – sujeito-verbo-objeto (SVO) – embora também se encontrem segmentos que constituem listas de enumerações, não possuindo, nesses casos, verbos, ou cujos verbos se encontram implícitos/não realizados. Em todos os casos, há uma proliferação de terminologia clínica.

O segundo documento para tradução, denominado “Textos Mais Frequentes” (daqui em diante, TMF), consiste num conjunto de textos pré-definidos passíveis de serem utilizados exclusivamente por profissionais de saúde (médicos e enfermeiros) para o registo de informação clínica sobre pacientes. Estes textos são provenientes de profissionais de saúde tanto dos EUA como do Reino Unido, pelo que apresentam ambas as variantes ortográficas do inglês (não foram detetadas variantes

terminológicas⁷⁹). Os textos que compõem o TMF não se cingem a uma única área de aplicação nem possuem uma relação direta entre si, na maior parte dos casos, contrariamente ao que acontece com os textos que compõem o IA. Os segmentos que compõem este ficheiro para tradução são títulos de categorias de informação, frases com espaços para preenchimento de informação por parte dos utilizadores ou frases com opções para seleção. Na sua maioria não seguem a estrutura sintática comum das frases em português (SVO) e em muitos casos são exclusivamente compostos por termos clínicos e não chegam a constituir frases completas.

A seleção dos textos para tradução teve como critério o facto de não terem tradução portuguesa associada (o que não significa, porém, que textos semelhantes não possam ter sido traduzidos no passado e, portanto, constem das memórias de tradução que constituem os *corpora* MMM utilizados neste estudo). Deste ficheiro foram eliminadas as repetições de forma automática no Excel, pois considerou-se que estas apenas iriam provocar “ruído”. No final, sobraram 96 segmentos para tradução, sendo que dois foram eliminados por não terem conteúdo a traduzir (um segmento em branco e outro apenas com reticências). No total ficaram 94 segmentos (correspondentes a 547 palavras) para tradução. Tal como no caso do primeiro *corpus* para tradução, também este foi segmentado através do SDL Trados Studio 2014 e, de seguida, transposto para o editor de texto Notepad++ com o título “textosmaisfrequentes.en”. A análise estatística no SDL Trados 2014⁸⁰ revelou que mais de metade dos segmentos a traduzir apresentam ocorrências entre 100% a 94% semelhantes a segmentos nas memórias de tradução que serviram de base aos *corpora* de treino usados no MMM.

4.3. Instrumentos de avaliação dos resultados da TA

As traduções realizadas pelo MMM e Google Tradutor foram avaliadas através de um sistema de avaliação automático (utilizando a métrica BLEU) e manual. Com a utilização de uma métrica automática pretendeu-se ter uma perspetiva mais imediata e geral da qualidade das traduções. Por outro lado, a introdução de uma avaliação manual teve como objetivo analisar em maior detalhe os problemas detetados nas traduções,

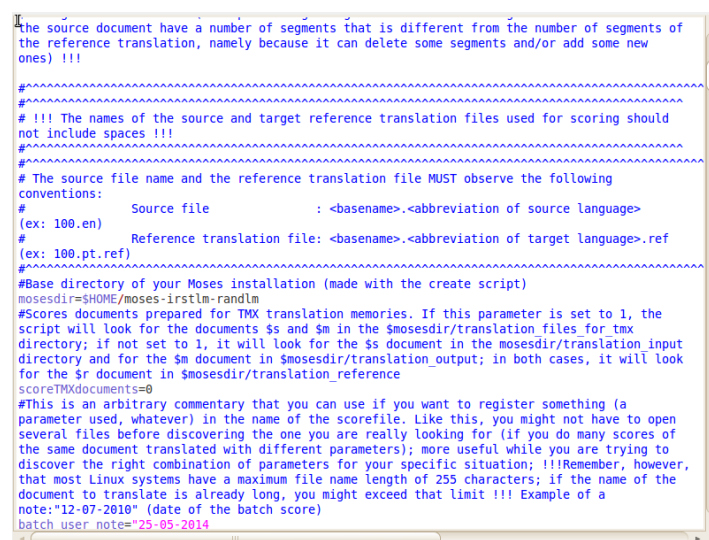
⁷⁹ Exemplo de variante terminológica: *Emergency Department*, no inglês dos EUA, e *Accident & Emergency*, no inglês do Reino Unido, para “Serviço de Urgência”, em português europeu.

⁸⁰ Vide Anexo xi, pp. 126

tanto em termos de tipo como de nível de gravidade, tendo como base critérios objetivos. Com estes dois instrumentos pretendeu-se ter uma visão complementar, ou seja, geral e pormenorizada.

4.3.1 Avaliação automática

A avaliação automática foi realizada através da utilização da métrica BLEU. O MMM possui um *script* que realiza uma avaliação BLEU, tendo por base uma tradução humana de referência. O mesmo *script* avalia ainda as traduções com a métrica NIST, contudo esta não foi utilizada para este projeto uma vez que se considerou que a complexidade introduzida pela adição de mais uma métrica automática não iria trazer informação suficientemente relevante que não fosse já traduzida pelos resultados da avaliação com a BLEU. As traduções de referência foram validadas por uma profissional clínica habilitada.



```
the source document have a number of segments that is different from the number of segments of
the reference translation, namely because it can delete some segments and/or add some new
ones) !!!

#-----
# !!! The names of the source and target reference translation files used for scoring should
not include spaces !!!
#-----
# The source file name and the reference translation file MUST observe the following
conventions:
#       Source file           : <basename>.<abbreviation of source language>
(ex: 100.en)
#       Reference translation file: <basename>.<abbreviation of target language>.ref
(ex: 100.pt.ref)
#-----
#Base directory of your Moses installation (made with the create script)
mosesdir=$HOME/moses-irstlm-randlm
#Scores documents prepared for TMX translation memories. If this parameter is set to 1, the
script will look for the documents $s and $m in the $mosesdir/translation files for tmx
directory; if not set to 1, it will look for the $s document in the $mosesdir/translation input
directory and for the $m document in $mosesdir/translation output; in both cases, it will look
for the $r document in $mosesdir/translation_reference
scoreTMXdocuments=0
#This is an arbitrary commentary that you can use if you want to register something (a
parameter used, whatever) in the name of the scorefile. Like this, you might not have to open
several files before discovering the one you are really looking for (if you do many scores of
the same document translated with different parameters); more useful while you are trying to
discover the right combination of parameters for your specific situation; !!!Remember, however,
that most Linux systems have a maximum file name length of 255 characters; if the name of the
document to translate is already long, you might exceed that limit !!! Example of a
note: "12-07-2010" (date of the batch score)
batch user note="25-05-2014"
```

Figura 22 – Script “score-0.85” do MMM.

A BLEU é uma das métricas mais conhecidas e difundidas. O conceito base desta métrica consiste em comparar n-gramas da TA com os n-gramas da tradução humana de referência e calcular o número de correspondências (Papineni, Roukos, Ward, & Zhu, 2002: 312). Quantas mais correspondências tiver, melhor será a tradução, aos olhos da BLEU. A este propósito atente-se nos exemplos de Papieni et al. (*ibidem*).

Candidata 1: *It is a guide to action which ensures that the military always obeys the commands of the party.*

Candidata 2: *It is to insure the troops forever hearing the activity guidebook that the party direct.*

Referência 1: *It is a guide to action that ensures that the military will forever heed the Party commands.*

Referência 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*

Referência 3: *It is the practical guide for the army always to heed the directions of the party.*

Neste caso, a frase candidata 1 seria considerada a melhor tradução por partilhar mais palavras e sintagmas com as três frases de referência (*ibidem*).

Um dos parâmetros fulcrais da BLEU é a métrica da precisão, segundo Koehn (2010: 226). A precisão, na BLEU, é calculada através do rácio de n-gramas corretos de uma determinada ordem *n* em relação ao número total de n-gramas gerados dessa ordem. Esta métrica de precisão é denominada precisão modificada de unigramas (*ibidem*). No caso do exemplo acima, a candidata 1 teria uma precisão modificada de unigrama de 17/18, enquanto a candidata 2 teria uma precisão de 8/14.

Um dos problemas das métricas relativas à precisão – a não penalização pela queda de palavras – é resolvido na BLEU através de uma penalização de brevidade. Esta penalização permite reduzir o resultado final caso a TA seja demasiado curta (*ibidem*).

Uma outra inovação da BLEU é a utilização de várias traduções de referência. Dada a variabilidade dos resultados de tradução, é difícil arranjar correspondências entre a TA e a tradução de referência. Contudo, se forem usadas várias traduções de referência, é mais provável que todas as traduções aceitáveis de partes ambíguas de frases sejam detetadas (*ibidem*: 227).

Estudos revelam existir uma correlação elevada entre os resultados da avaliação humana e da avaliação automática, ou seja, traduções avaliadas com pontuação elevada por métricas automáticas são normalmente avaliadas com pontuação igualmente elevada por revisores humanos e vice-versa. Apesar de imperfeitas e com grande espaço para melhorias, as métricas automáticas são extremamente úteis para fornecer informação relativamente fidedigna e de forma rápida acerca da qualidade de uma tradução.

4.3.2. Avaliação manual

Para a avaliação manual, optámos por criar um sistema de avaliação manual próprio, já que os sistemas de avaliação manual da qualidade da tradução disponíveis não nos pareceram adequados ao objetivo deste estudo de caso. Por um lado, os sistemas de avaliação da qualidade da tradução humana mais utilizados no mercado da tradução (como, por exemplo, SAE J2450, LISA QA Model, ou o SDL Classic Model) não apresentavam os critérios que desejávamos avaliar neste estudo de caso; por outro, várias ferramentas de avaliação manual da tradução automática eram demasiado complexas de instalar e de operar⁸¹.

O sistema de avaliação manual desenvolvido para este projeto⁸² tem por base uma lista de categorias de erros/problemas detetados, identificados por um código de cores, e classificados de acordo com a sua gravidade.

As categorias de erros consideradas foram as seguintes:

Tabela 1 – Categorias de problemas detetados na tradução.

Gramática	Problemas de ordem sintática e/ou semântica, incluindo erros relacionados com a ordem das palavras, concordância, seleção do tempo e modo verbal, seleção de casos e marcas diacríticas.
Terminologia	Sempre que o termo utilizado não é o mais apropriado tendo em conta o contexto e o domínio; ou sempre que não é utilizado um termo do mesmo domínio; ou o termo utilizado não corresponde ao mesmo conceito no domínio do TP.
Pontuação	É considerado erro de pontuação sempre que as convenções da pontuação da LC não forem tidas em consideração ou adulterarem a mensagem do TC em relação ao TP.
Omissão/adição	Este erro ocorre sempre que alguma informação do TP seja

⁸¹ Vide, por exemplo, os programas de avaliação manual disponíveis na Internet, como o Appraise (Federman, 2012) e Costa MT Tool (Chatzitheodorou e Chatzistamatis 2010).

⁸² O sistema de avaliação manual criado foi bastante inspirado nos critérios de avaliação das provas da ATA – American Translators Association.

	omitida/adicionada injustificadamente no TC.
Erro de tradução	Existe um erro de tradução sempre que a mensagem de um segmento no TP não for devidamente transferida para o TC (inclui seleção lexical).
Má utilização de maiúsculas e minúsculas	Sempre que não forem mantidas as maiúsculas/minúsculas conforme o TP, exceto quando a convenção que rege a utilização de maiúsculas/minúsculas no TC se sobrepuser à utilização que é feita das maiúsculas/minúsculas no TP.
Coerência	Incoerência na utilização de terminologia ou na tradução de léxico geral ao longo do TC.

Todos os segmentos foram analisados e os problemas detetados foram devidamente assinalados com a cor correspondente. Aos problemas detetados foi atribuída uma classificação de gravidade relativa ao grau de impacto destes na tradução. A classificação do nível de impacto do problema no TC foi a seguinte:

Tabela 2 – Níveis de gravidade/impacto de problemas detetados na tradução.

1	Ligeiro	Não interfere com a interpretação do segmento do TC em causa nem adultera a mensagem em relação ao TP.
2	Moderado	Interfere com a interpretação do segmento do TC em causa e adultera parte da mensagem em relação ao TP.
3	Grave	Impossibilita ou adultera de modo irreparável a mensagem transmitida no segmento do TC em relação ao TP.

Depois de processados no MMM e no Google Tradutor, tanto o texto original em inglês como a respetiva tradução foram transferidos para um documento Word. Ao texto original e traduzido, devidamente segmentados, foi anexada uma tabela com uma coluna para cada uma das categorias de erros/problemas descritos acima. Cada segmento traduzido foi analisado e os problemas detetados foram assinalados com a cor

da categoria correspondente. Além de categorizados, os problemas foram ainda classificados numa escala de 1 a 3, conforme o impacto do erro no segmento traduzido.

	ORIGINAL_EN	MMM_PT (ALLALLEMEA)	Gra	Ter	Po	Om	Ta	Mai	Coe
1	Chronic hypertension	Crónica hipertensão	1						
2	Chronic hypertension is a long-term condition in which your blood pressure (BP) is higher than normal.	Crónica a hipertensão é uma doença de longa duração em que a tensão arterial (PA) é superior à normal.	1						1
3	Your BP is the force of your blood moving against the walls of your arteries.	Sua tensão arterial esta a força da movimentação contra as paredes das suas artérias.				2	1		1
4	Normal blood pressure is less than 120/80.	Sua tensão arterial for inferior a 120/80.				2	1		1
5	Prehypertension is BP between 120/80 and 139/89.	Prehypertension é entre PA 120/80 e 151/93.	1				1		

Figura 23 – Exemplo de tabela usada para a análise, catalogação e classificação dos problemas detetados em cada segmento.

Quando mais do que um erro foi detetado numa única palavra, esta foi assinalada com as cores correspondentes aos problemas identificados.

24	These arteries carry blood from your heart to your lungs to pick up oxygen.	Essas artérias Pegue sangue do seu coração para os pulmões pegar oxigénio.						3	1
----	---	--	--	--	--	--	--	---	---

Figura 24 – Exemplo de duas categorias de problemas detetados numa única palavra.

Quando um segmento apresentou mais do que um problema da mesma categoria, a classificação atribuída teve em consideração o impacto do agregado desses erros no segmento.

30	The lower right chamber of your heart, the right ventricle, pumps blood to your pulmonary arteries.	Canto inferior direito da câmara do coração, do ventrículo direito, aspiradores de sangue para a tua pulmonares artérias.	2				3		
----	---	---	---	--	--	--	---	--	--

Figura 25 – Exemplo de atribuição de nível de gravidade quando a mesma categoria de problema ocorre mais do que uma vez.

A avaliação foi realizada por uma única pessoa.

5. Resultados

Neste capítulo iremos apresentar os resultados da análise à tradução dos textos “Instruções de alta” (IA) e “Textos Mais Frequentes” (TMF) realizada pelos sistemas MMM e Google Tradutor. Para facilitar a compreensão, o *corpus* CorpusALLALL daqui em diante passa a ser referido simplesmente por “ALL”; o CorpusALLALLEMEA passa a ser apenas “EMEA”; o *corpus* CorpusALLALLEMEAEuroparl7 passa a “Europarl”; e Google Tradutor passa apenas a ser “Google”.

A tradução de ambos os documentos – IA e TMF – através de diferentes “motores” MMM, treinados a partir de três tipos de *corpora* diferentes, conduziu a resultados bastante díspares em certas alturas, mas também a resultados bastante semelhantes noutras.

5.1. Resultados da avaliação manual

5.1.1. Resultados da tradução de “Instruções de Alta”

Em termos gerais, qualquer uma das traduções realizadas pelo MMM apresentou maior ocorrência de problemas em relação à tradução realizada pelo Google.

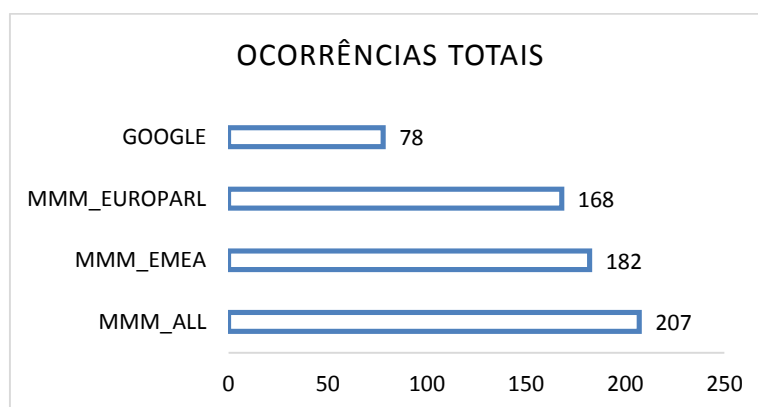


Figura 26 – Número de ocorrências totais de problemas nas traduções de “Instruções de alta” realizadas pelo MMM e Google Tradutor.

Conforme é possível verificar na Figura 27, as traduções com maior ocorrência de problemas foram as realizadas pelo MMM: ALL (207), EMEA (182) e Europarl (168), respetivamente. A tradução do Google (78) foi a que apresentou menos ocorrências.

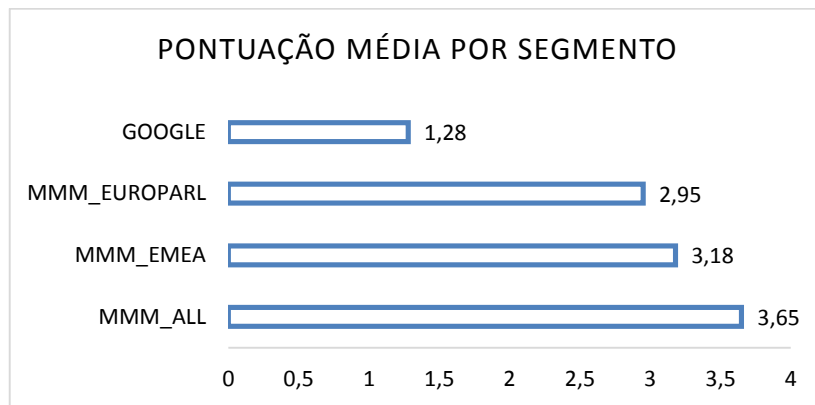


Figura 27 – Pontuação média por segmento nas traduções de "Instruções de alta" realizadas pelo MMM e Google Tradutor.

Quanto à pontuação média por segmento (Figura 28), a tradução produzida pelo Google foi aquela que obteve a melhor média (1,28), seguida do Europarl (2,95). A tradução que teve por base o EMEA (3,18) surge na terceira posição e, finalmente, a do ALL (3,65) é a que apresenta pior qualidade média por segmento.

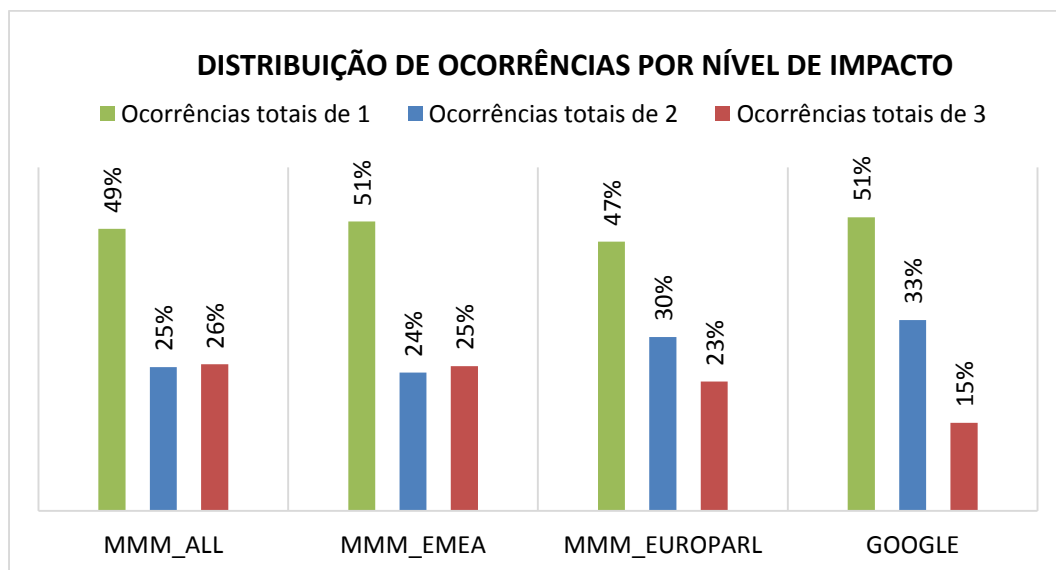


Figura 28 – Percentagem de ocorrências por nível de impacto nas traduções de "Instruções de alta" realizadas pelo MMM e Google Tradutor.

Em termos de distribuição de problemas por nível de gravidade (Figura 29), todas as traduções revelaram uma predominância de erros de nível 1: ALL – 49%, EMEA – 51%, Europarl – 47% e Google – 51%. No que diz respeito a erros de gravidade moderada e grave, a percentagem de problemas de nível 2 detetados nas traduções do

Europarl e Google superaram claramente a percentagem de problemas de nível 3 (Europarl – 30% / 23% e Google – 33% / 15%). Por outro lado, nas traduções do ALL e do EMEA, as ocorrências registadas de nível 2 e 3 foram praticamente na mesma proporção: ALL – 25% / 26% e EMEA – 24% / 25% .

Em termos da distribuição dos erros detetados por categoria e respetivo nível de impacto, a Figura 30 apresenta os resultados apurados para cada uma das traduções.

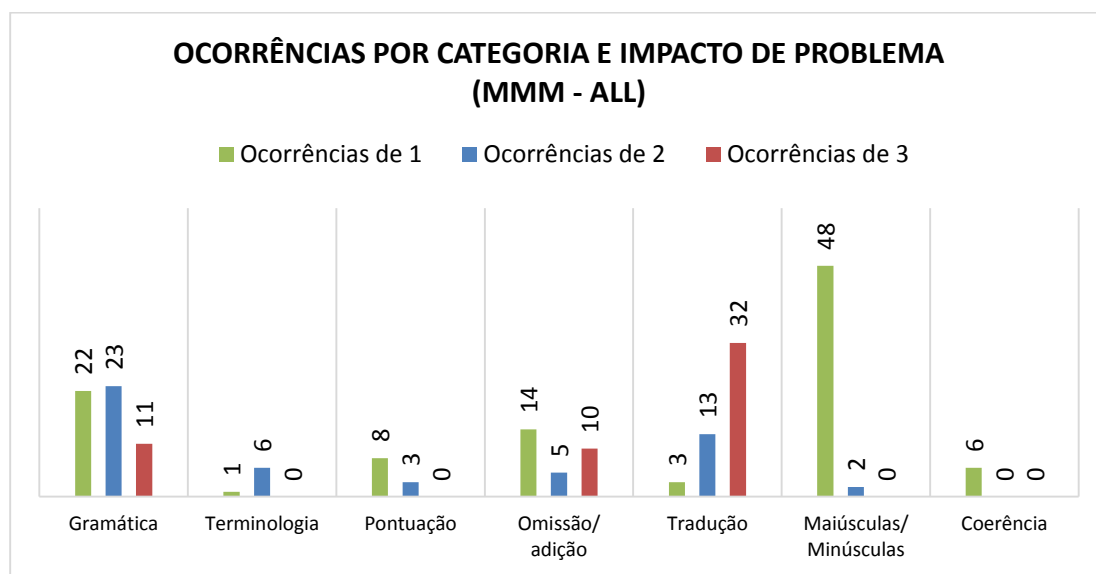


Figura 29 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Instruções de alta" realizada pelo MMM com o corpus ALL.

Através da análise à Figura 30 verifica-se que a maioria dos problemas registados na tradução do ALL se encontrava relacionada com “Gramática” (56 ocorrências no total), “Maiúsculas/minúsculas” (50 ocorrências no total), “Tradução” (48 ocorrências no total) e “Omissão/adição” (29 ocorrências no total). “Coerência” (6), “Terminologia” (7) e Pontuação (11) foram as categorias de problemas com menos ocorrências. A categoria com mais problemas de nível 1 foi “Maiúsculas/minúsculas” com 48 ocorrências, seguida de “Gramática” (22), “Omissão/adição” (14), “Pontuação” (8), “Coerência” (6), “Tradução” (3) e “Terminologia” (1). Em termos de problemas de gravidade de nível 2, as categorias com maior número de ocorrências foram “Gramática” (22) e “Tradução” (13). “Terminologia” (6), “Omissão/adição” (5), “Pontuação” (3) e “Maiúsculas/minúsculas” (2) foram as categorias com menos ocorrências de problemas com este tipo de gravidade. Em termos de “Coerência” não foi registado qualquer problema de nível 2. Quanto a problemas mais graves, ou seja, de

nível 3, estes concentraram-se apenas em três categorias: “Tradução” (32), “Gramática” (11) e “Omissão/adição” (10). Em nenhuma outra categoria houve qualquer registo de problemas deste nível.

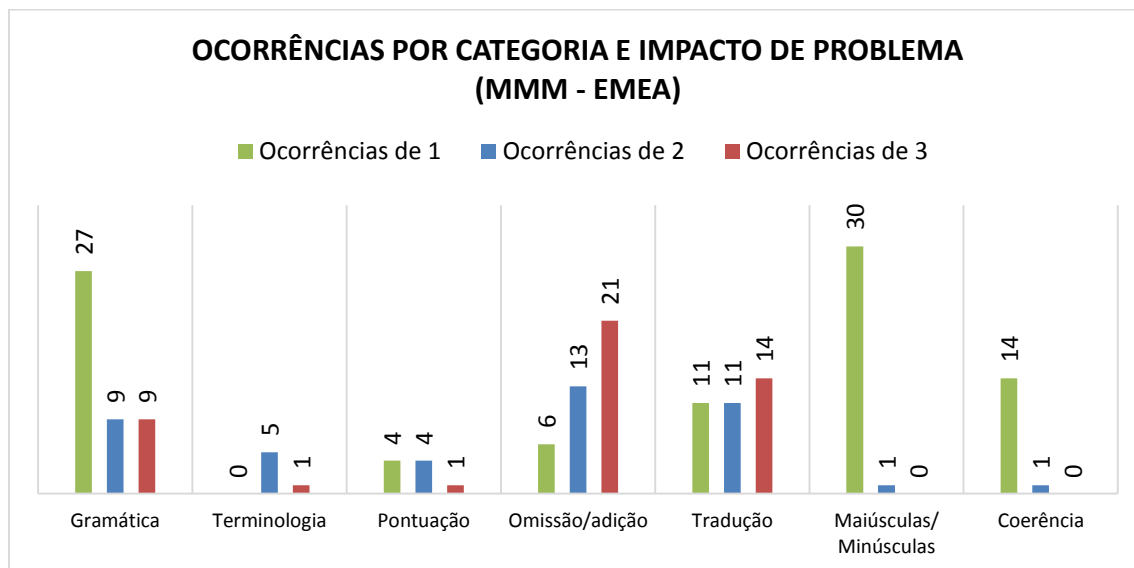


Figura 30 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Instruções de alta" realizada pelo MMM com o corpus EMEA.

Quanto à tradução produzida pelo MMM tendo por base o *corpus* EMEA, é possível verificar, através da Figura 31, que as categorias onde se registaram mais problemas foram: “Gramática” (45 no total), “Omissão/adição” (40 no total) e “Tradução” (36 no total). Em termos de “Gramática”, as ocorrências foram maioritariamente de impacto ligeiro (27), embora registre 9 ocorrências tanto de impacto moderado como de grave. Em termos de ocorrências graves, a categoria com o maior número de problemas foi a “Omissão/adição”, com 21 ocorrências, seguido de “Tradução”, com 14. Também a “Terminologia” e a “Pontuação” registaram erros graves, mas apenas 1 cada. Quanto a ocorrências de impacto moderado, foi igualmente à “Omissão/adição” que pertenceu o maior número de ocorrências, com 13, seguida novamente pela “Tradução”, com 11. As restantes categorias não ultrapassaram as 10 ocorrências: “Gramática” – 9, “Terminologia” – 5, “Pontuação” – 4, e “Maiúscula/minúscula” e “Coerência” – 1 cada. No que diz respeito a falhas de nível ligeiro, o destaque foi para a categoria “Maiúsculas/minúsculas”, com 30 ocorrências, seguida da “Gramática” com 27, “Coerência”, com 14, “Tradução”, com 11, “Omissão/adição” com 6 e “Pontuação” com 4. A “Terminologia” não registou qualquer ocorrência de nível 1.

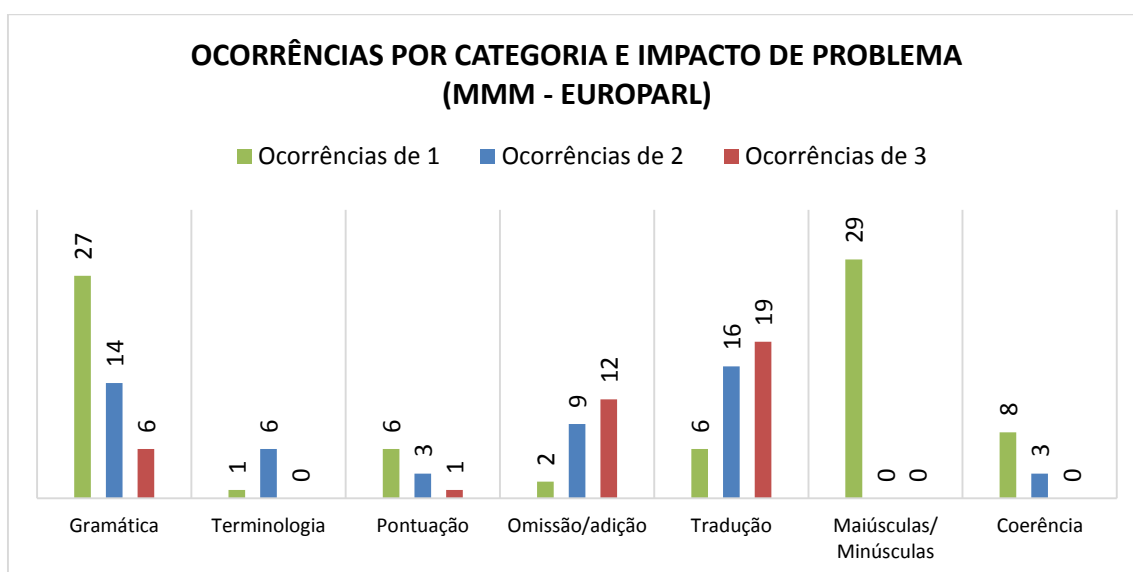


Figura 31 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Instruções de alta" realizada pelo MMM com o corpus Europarl.

No que diz respeito à tradução baseada no *corpus* Europarl, do MMM, a Figura 32 revela que as ocorrências se concentraram principalmente nas áreas da “Gramática” (47 no total), “Tradução” (41 no total) e “Maiúsculas/minúsculas” (29), embora neste último caso, sejam todas de impacto mais baixo. As ocorrências mais graves, contudo, concentraram-se principalmente nas categorias “Tradução” (19) e “Omissão/adição” (12), embora também a “Gramática” (6) e a “Pontuação” (1) registassem ocorrências com este tipo de impacto. Quanto a ocorrências de nível intermédio, estas concentraram-se na “Tradução” (16), “Gramática” (14) e “Omissão/adição” (9). As restantes categorias apresentaram menos ocorrências a este nível: “Terminologia”, 6, e “Pontuação” e “Coerência”, ambas registaram 3 cada. Em termos de ocorrências de impacto baixo, além das “Maiúsculas/minúsculas” (29), também a “Gramática” registou bastantes falhas deste nível (27), seguindo-se a “Coerência” (8), “Pontuação” e “Tradução” (6 cada), “Omissão/adição” (2) e “Terminologia” (1).

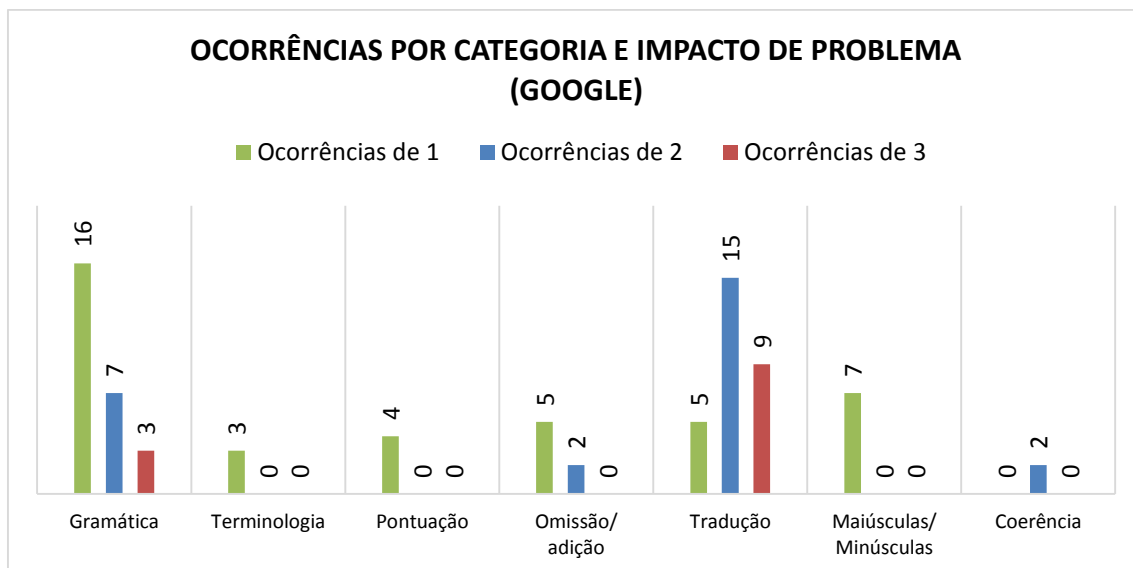


Figura 32 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Instruções de alta" realizada pelo Google Tradutor.

Quanto à tradução realizada pelo Google, o maior número de ocorrências concentrou-se nas áreas de “Tradução” (29 no total) e “Gramática” (26 no total), tal como é possível verificar na Figura 33. Foram igualmente estas categorias as únicas a registar falhas de impacto grave – “Tradução”, 9; e “Gramática”, 3 – e a registar o maior número de ocorrências de impacto moderado (15 e 7, respetivamente). As restantes categorias a registar problemas deste nível foram a “Omissão/adição” e a “Coerência”, com 2 ocorrências cada. À exceção da “Coerência”, todas as outras categorias registam ocorrências de nível 1: “Gramática” (16), “Maiúsculas/minúsculas” (7), “Omissão/adição” e “Tradução” (5), “Pontuação” (4) e “Terminologia” (3).

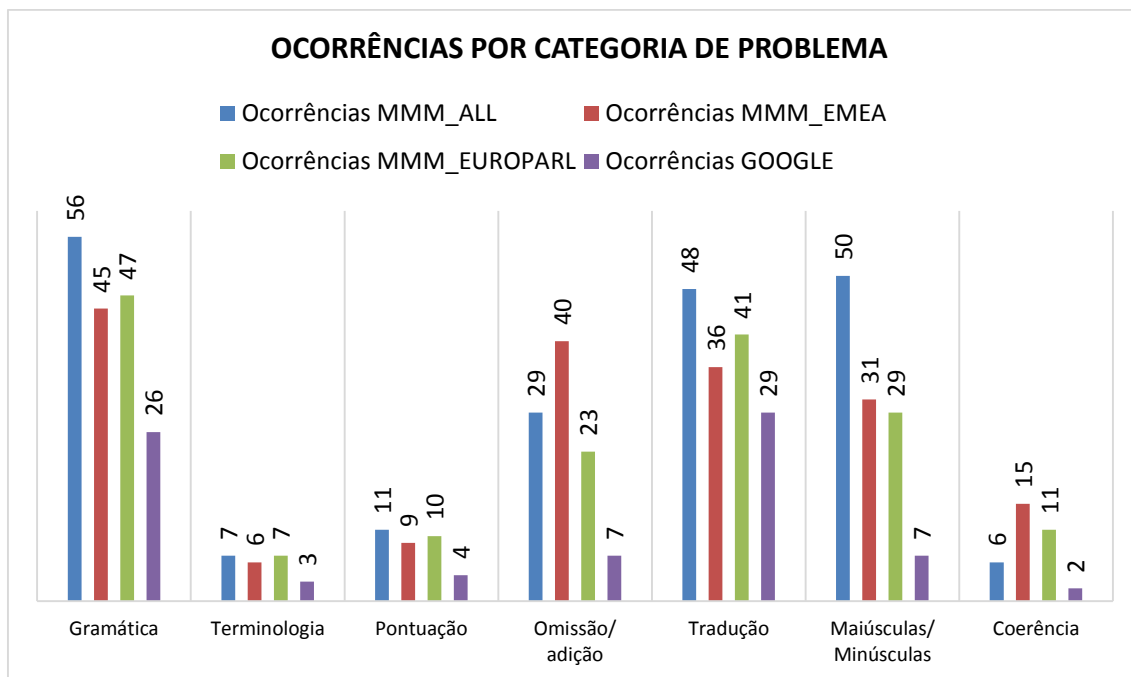


Figura 33 – Ocorrências por categoria de problema na tradução de "Instruções de alta" realizada pelo MMM e Google Tradutor.

Numa visão de agregado, como aquela que é dada através da Figura 34, é possível verificar que a maioria dos problemas detetados nas traduções produzidas pelos dois sistemas ocorreu ao nível da “Gramática” (174 no total), “Tradução” (154 no total), “Maiúsculas/minúsculas” (117 no total) e “Omissão/adição” (99 no total). A tradução do Google foi aquela que registou menos ocorrências em todas as categorias (cf. Tabela 2). Por outro lado, a tradução realizada pelo MMM tendo por base o *corpus* ALL foi aquela que apresentou maior número de ocorrências em quase todas as categorias, exceto na “Omissão/adição” (29) e “Coerência” (6), em que foi ultrapassada pela tradução do EMEA (40 e 15, respetivamente)

5.1.2. Resultados da tradução de “Textos Mais Frequentes”

Em termos gerais, as traduções realizadas pelo MMM apresentaram maior número de ocorrências em relação à tradução realizada pelo Google, conforme é possível verificar através da Figura 35.

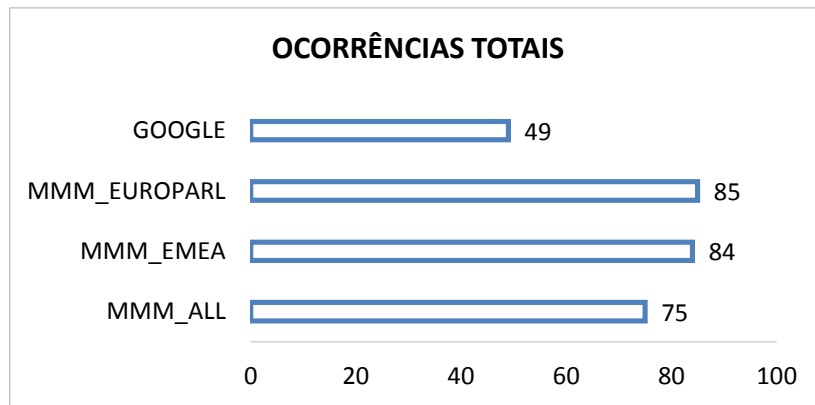


Figura 34 – Número de ocorrências totais de problemas nas traduções de “Textos mais frequentes” realizadas pelo MMM e Google Tradutor.

A tradução realizada pelo Google foi aquela que apresentou menor número de problemas detetados (49), em relação às traduções produzidas pelo MMM. A tradução produzida pelo MMM baseada no *corpus* ALL foi a que apresentou menos ocorrências, entre as traduções realizadas pelo MMM, com 75 problemas detetados. Esta foi seguida pelas traduções do EMEA (84) e Europarl (85).

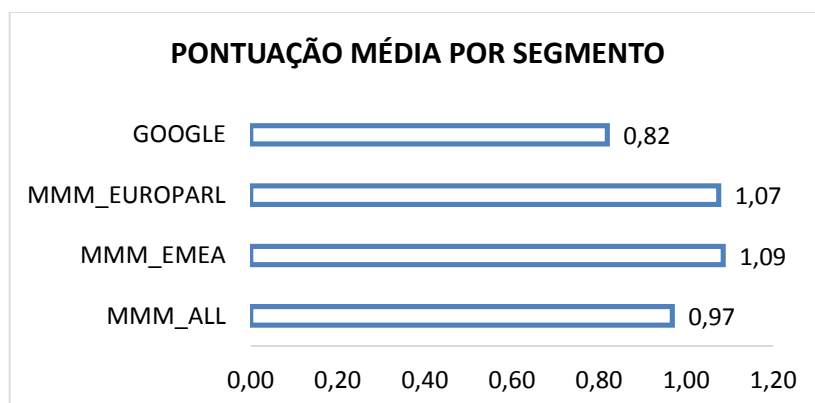


Figura 35 – Pontuação média por segmento das traduções de “Textos mais frequentes” realizadas pelo MMM e Google Tradutor.

No que diz respeito à pontuação média por segmento (Figura 36), a tradução do Google foi aquela que obteve melhores resultados, com 0,82, seguida da tradução baseada no *corpus* ALL, com 0,97. As traduções realizadas pelo MMM com base no Europarl e EMEA apresentaram uma pontuação muito semelhante: 1,07 e 1,09, respetivamente.

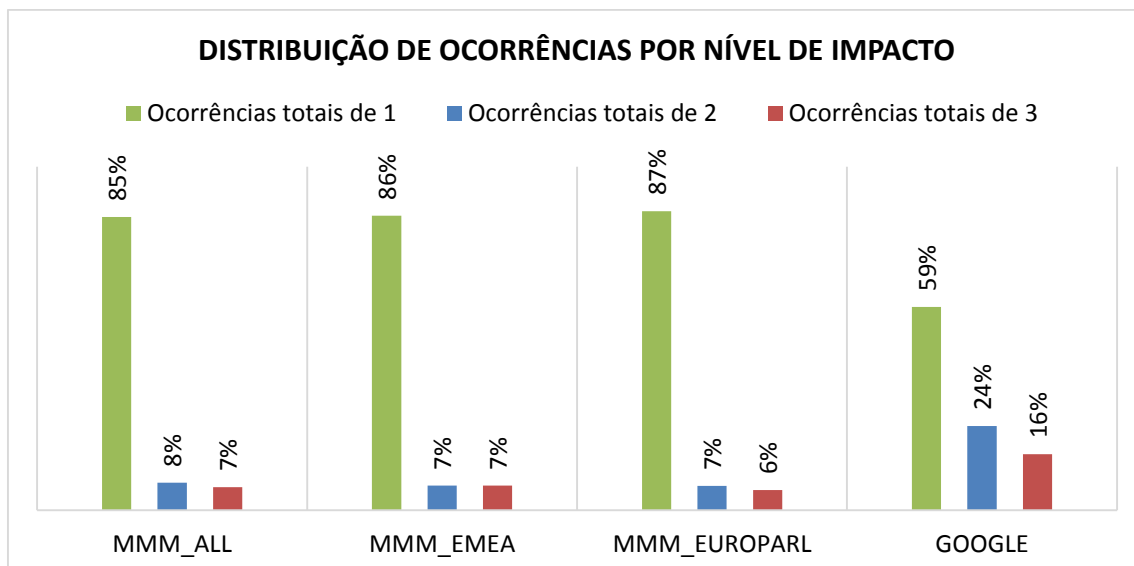


Figura 36 – Percentagem de ocorrências por nível de impacto na tradução de “Textos mais frequentes” realizada pelo MMM e Google Tradutor.

Relativamente à distribuição dos problemas detetados por nível de gravidade, a Figura 37 revela que em todas as traduções predominaram as falhas de gravidade mais baixa, ou seja, de nível 1. A percentagem de falhas deste nível foi mais elevada nas traduções realizadas pelo MMM (ALL – 85%, EMEA – 86% e Europarl – 87%), em relação à tradução do Google (59%). Em relação aos restantes níveis de gravidade, as traduções produzidas pelo MMM apresentaram uma percentagem consideravelmente menor do que o Google: enquanto 24% dos problemas detetados na tradução do Google foram de gravidade moderada, no caso das traduções do MMM a percentagem mais elevada não ultrapassou os 8% (ALL – 8%, EMEA – 7% e Europarl – 7%); em relação a falhas graves (nível 3), 16% das falhas registadas na tradução do Google foram desse nível, enquanto nas traduções do MMM a percentagem não excede os 7% (ALL – 7%, EMEA – 7% e Europarl – 6%).

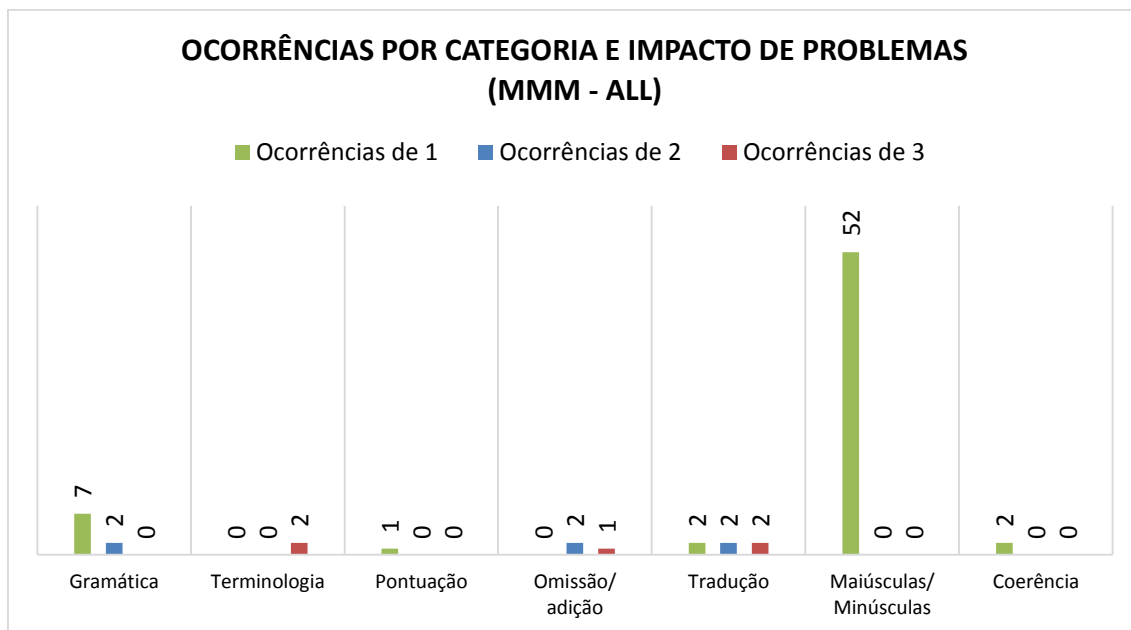


Figura 37 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes" realizada pelo MMM com o *corpus* ALL.

Ao nível da distribuição dos problemas detetados em cada uma das traduções foi possível verificar, através da Figura 38, que, no que diz respeito à tradução realizada pelo ALL, a esmagadora maioria dos problemas detetados eram relacionados com “Maiúsculas/minúsculas” (52). A segunda categoria com maior número de falhas foi a “Gramática”, com 9 falhas detetadas, 7 de nível ligeiro e 2 de nível moderado. A categoria “Tradução” apresentou 6 falhas, 2 de cada nível (ligeiro, moderado e grave). Além da “Tradução”, as únicas categorias com registo de problemas graves foram a “Terminologia”, com 2 ocorrências, e a “Omissão/adição” com 1 (esta última apresentou ainda 2 ocorrências de nível 2). Tanto a “Coerência” como a “Pontuação” registaram apenas ocorrências de nível 1: “Coerência” – 2 e “Pontuação” – 1.

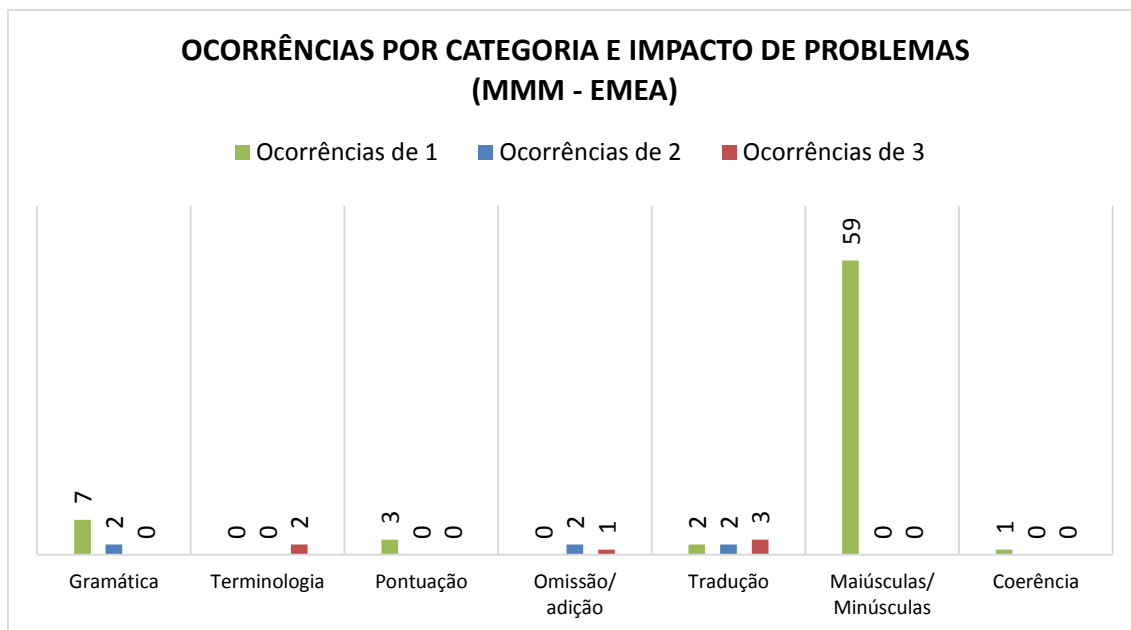


Figura 38 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes" realizada pelo MMM com o *corpus* EMEA.

Tal como na tradução baseada no *corpus* ALL, também na do EMEA a grande maioria dos problemas detetados esteve relacionada com “Maiúsculas/minúsculas” (59), conforme é possível verificar na Figura 39. Da mesma forma, a segunda categoria com maior número de falhas foi a “Gramática”, com 9 ocorrências, 7 de gravidade ligeira e 2 de gravidade intermédia. A categoria “Tradução” apresentou 7 ocorrências no total: 2 de nível ligeiro, 2 de nível moderado e 3 de nível grave. Além da “Tradução”, as únicas categorias com registo de problemas graves foi a “Terminologia”, com 2 ocorrências, e a “Omissão/adição” com 1 (esta última apresenta ainda 2 ocorrências de nível moderado). Tanto a “Coerência” como a “Pontuação” registaram apenas ocorrências de nível ligeiro: “Pontuação”, 3; e “Coerência”, 1.

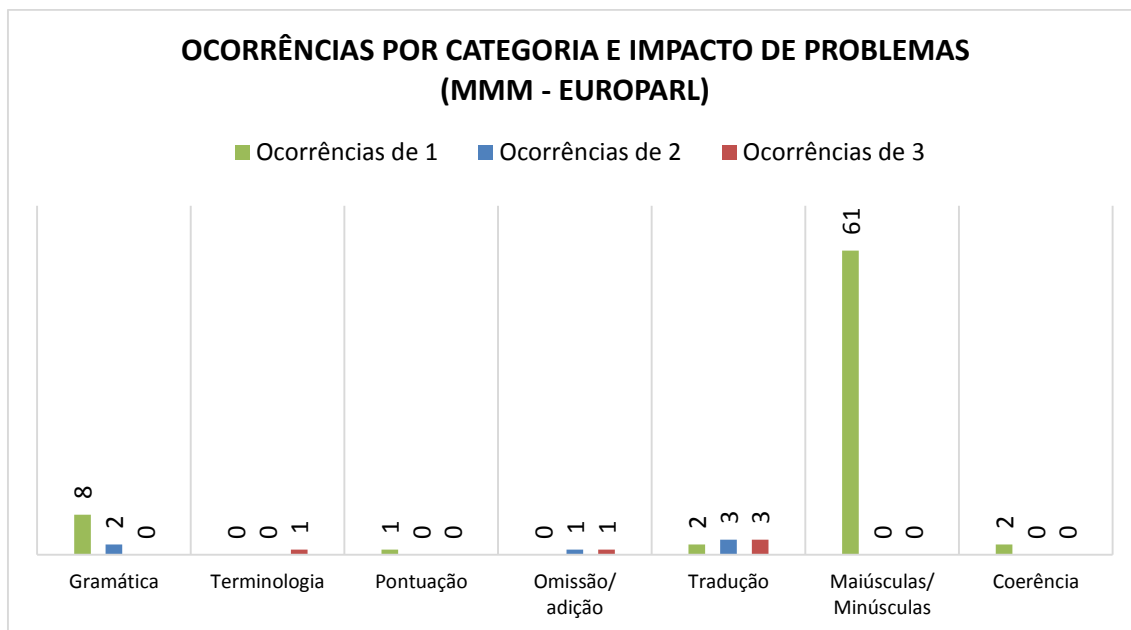


Figura 39 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes" realizada pelo MMM com o *corpus* Europarl.

Da mesma forma que nas traduções baseadas nos *corpora* ALL e EMEA, também na do Europarl a grande maioria dos erros detetados foi relacionada com “Maiúsculas/minúsculas” (61), conforme é possível verificar na Figura 40. A segunda categoria com maior número de problemas detetados foi a “Gramática”, com 10 ocorrências, 8 de gravidade ligeira e 2 de gravidade moderada (nenhuma de gravidade maior). A categoria “Tradução” apresentou 8 ocorrências no total: 2 de nível mais baixo e 3 de nível moderado e grave. Além da “Tradução”, as únicas categorias com registo de problemas graves foram a “Terminologia” e a “Omissão/adição” cada uma com 1 ocorrência. A “Omissão/adição” apresentou ainda 1 ocorrência de nível moderado, ao contrário da “Terminologia” que não apresentou qualquer outro tipo de falha. A “Coerência” apresentou 2 falhas de nível baixo e a “Pontuação” registou apenas 1 ocorrência do mesmo nível.

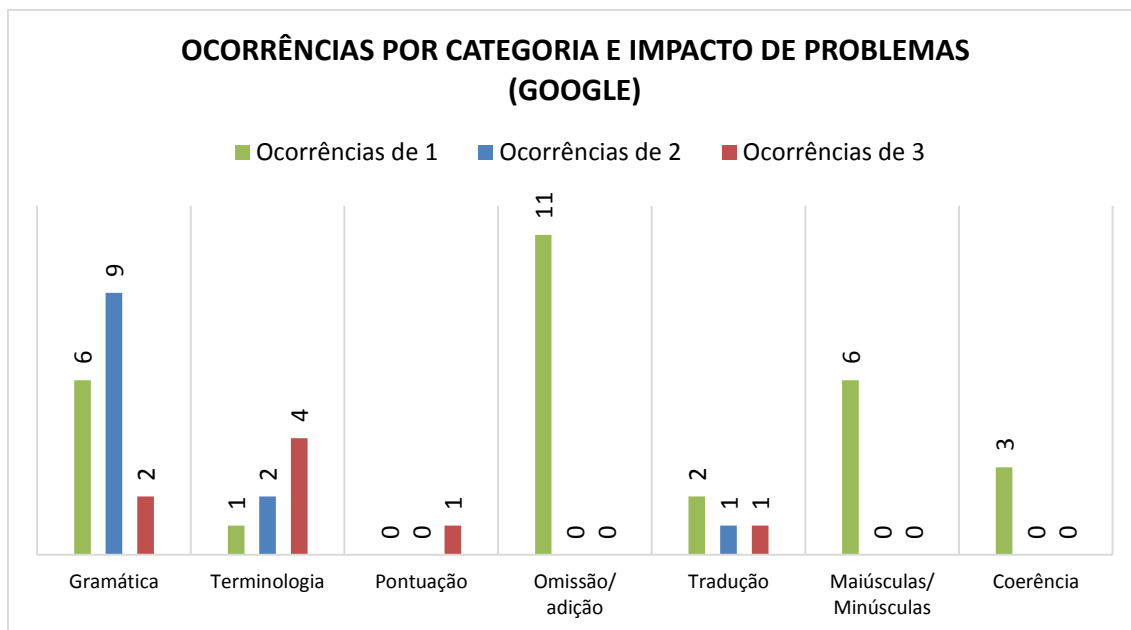


Figura 40 – Distribuição da ocorrência de problemas por categoria e impacto na tradução de "Textos mais frequentes" realizada pelo Google.

No caso da tradução realizada pelo Google, as ocorrências de problemas detetados encontraram-se mais distribuídas por todas as categorias de erros (Figura 41). A categoria com maior número de falhas foi a “Gramática”, com 17 ocorrências, 6 de nível ligeiro, 9 de nível moderado e 2 de nível grave. A segunda categoria com mais erros detetados foi a “Omissão/adição”, mas, neste caso, todas as 11 ocorrências são de nível baixo. A “Terminologia” foi aquela que apresentou maior número de ocorrências graves (4). Além desta, também a “Pontuação”, com 1 ocorrência registada, e a “Tradução”, também com 1 ocorrência, registaram erros de nível grave. Tanto a categoria “Maiúsculas/minúsculas” como a “Coerência” apresentaram apenas ocorrências de nível baixo (6 e 3, respetivamente).

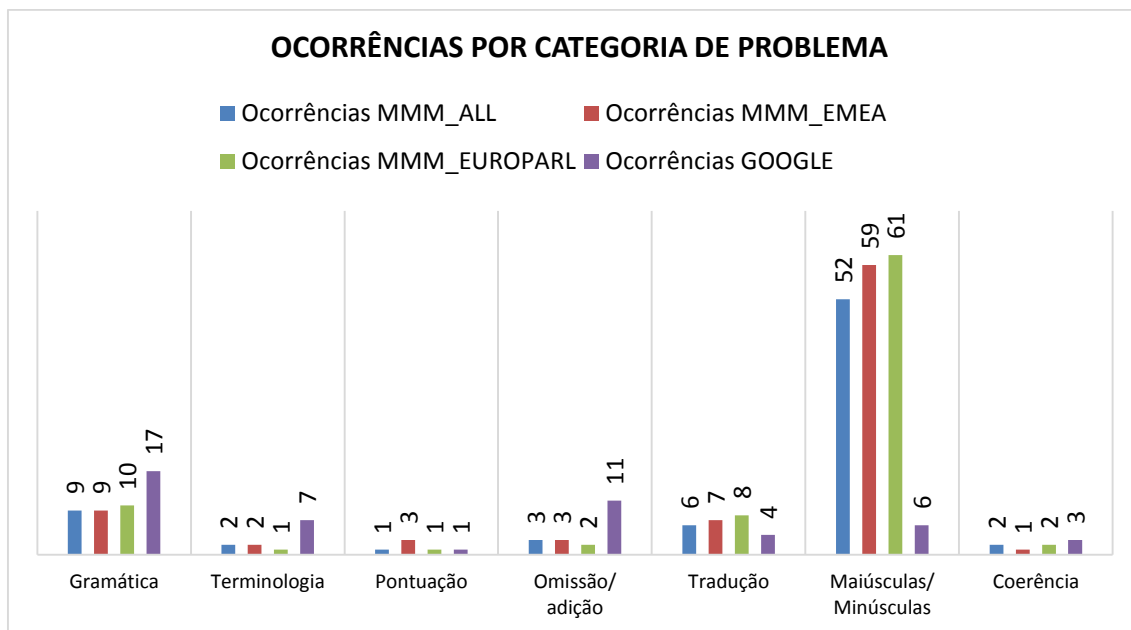


Figura 41 – Ocorrências por categoria de problema nas traduções de "Textos mais frequentes" realizadas pelo MMM e Google Tradutor.

A Figura 42 revela que as áreas em que, tanto as traduções do MMM como do Google, apresentaram maior número de problemas foram as áreas “Maiúsculas/minúsculas” (com 178 ocorrências no total), “Gramática” (com 45 ocorrências no total) e “Tradução” (com 25 ocorrências no total). As traduções do MMM apenas registaram mais ocorrências do que a do Google nas categorias “Pontuação” (EMEA – 3, ALL, Europarl e Google – 1), “Tradução” (ALL – 6, EMEA – 7, Europarl – 8 e Google – 4) e “Maiúsculas/minúsculas” (ALL – 52, EMEA – 59 e Europarl – 61, em comparação com 6 do Google). Em todas as restantes categorias, a tradução do Google registou mais ocorrências do que as do MMM, ou seja, em, “Gramática” (17), “Terminologia” (7), “Omissão/adição” (11) e “Coerência” (3).

5.1.3 Comparativo entre resultados da tradução de “Instruções de Alta” e “Textos Mais Frequentes”

Analisando a prestação de MMM e Google de acordo com o tipo de documento traduzido, é possível verificar através da Figura 43, que ambos os sistemas tiveram mais dificuldades em traduzir o documento IA do que o TMF.

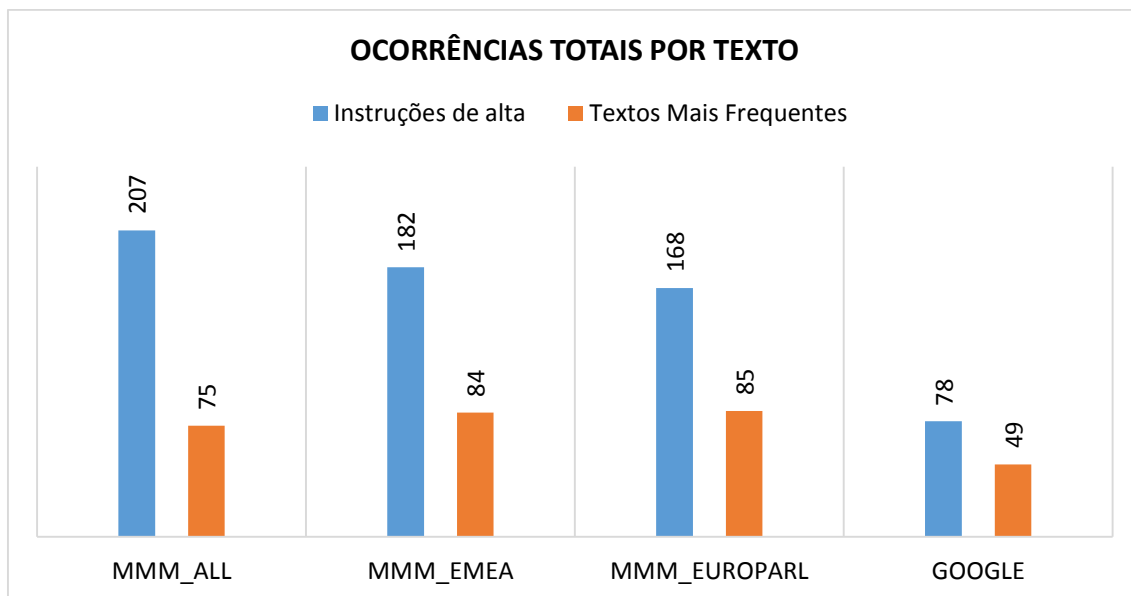


Figura 42 – Número de ocorrências totais de problemas dos textos traduzidos pelo MMM e Google Tradutor.

Esta diferença foi mais acentuada nas traduções realizadas pelo MMM do que nas do Google. Nas traduções produzidas pelo Google não houve uma diferença muito marcada em termos de número de problemas detetados entre o IA e o TMF (78/49, respetivamente), ao contrário do que aconteceu nas traduções produzidas pelo MMM (ALL – 207/75, EMEA – 182/84 e Europarl – 168/85).

No caso da tradução de IA, o número de ocorrências diminuiu de acordo com a dimensão do *corpus*, ou seja, quanto maior era o *corpus* de treino do sistema de TAE, menor foi o número de ocorrências verificado: ALL – 207, EMEA – 182, Europarl – 168, Google – 49.

O comportamento dos sistemas na tradução de TMF foi bastante diferente. Apesar de a tradução realizada pelo Google ter sido aquela que apresentou menos ocorrências dentre todas, a tradução realizada com base no ALL foi aquela que apresentou o menor número de ocorrências (75) entre as traduções realizadas pelo MMM, tendo sido seguida pela tradução do EMEA (84) e só depois pela do Europarl (85) (Figura 43).

É ainda relevante notar que a diferença de ocorrências entre a tradução do IA e TMF é muito maior entre motores MMM (ALL – 207/75, EMEA – 182/84 e Europarl – 168/85) do que no Google (78/49).

Contudo, se se excluir a contabilização de problemas relacionados com “Maiúsculas/minúsculas” (daqui em diante “M/M”), visto que, nas línguas em apreço (português e inglês), a utilização das maiúsculas não tem impacto nos julgamentos de

gramaticalidade de uma frase nem acarreta problemas de tradução, verificam-se alterações significativas nos resultados.

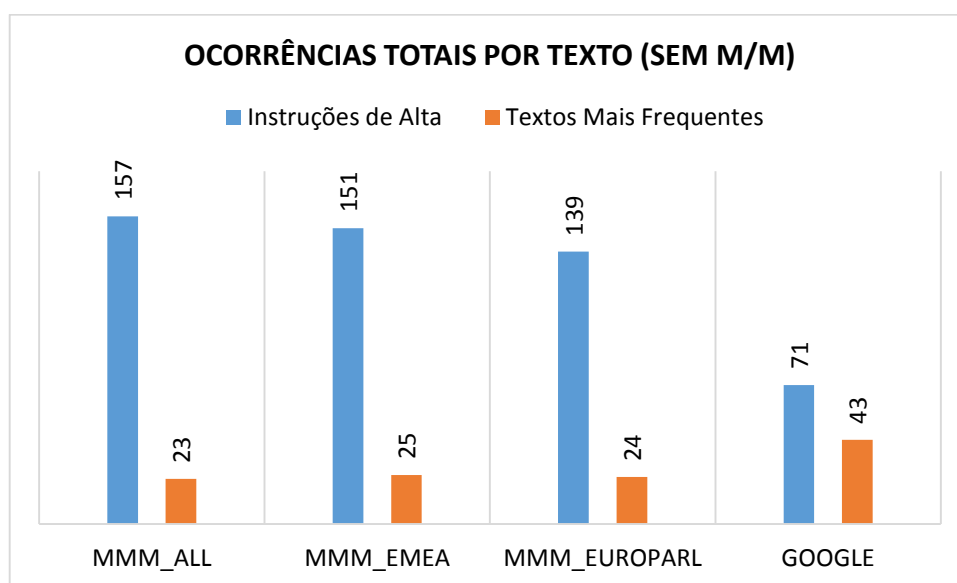


Figura 43 – Número de ocorrências totais de problemas (sem contabilizar “M/M”) dos textos traduzidos pelo MMM e Google Tradutor.

De acordo com a Figura 44, apesar de as traduções de IA apresentarem uma tendência semelhante, ou seja, de diminuição do número de problemas registados de acordo com o aumento da dimensão do *corpus* de treino (ALL – 157, EMEA – 151, Europarl – 139 e Google – 71), o mesmo não acontece nas traduções de TMF. Se com “M/M” a tradução do Google foi aquela que apresentou menos problemas registados em comparação com as traduções do MM, sem “M/M” a tradução do Google foi a que apresentou mais ocorrências (Google – 43, Europarl – 24, EMEA – 25 e ALL – 23). A tradução do ALL foi aquela que apresentou menos registos de problemas. Aliás, os resultados de todas as traduções do MMM, sem “M/M”, foram muito semelhantes.

Com a exclusão de “M/M” tornou-se ainda mais marcada a discrepância entre o número de ocorrências registadas entre as traduções de IA e TMF nos motores MMM (ALL – 157/23, EMEA – 151/25 e Europarl – 139/24), o que não aconteceu com o Google (71/43).

Em termos de pontuação média por segmento (Figura 45), ou seja, a medida utilizada para obter uma noção imediata e muito genérica acerca da qualidade de tradução⁸³, os

⁸³ Quanto menor for a pontuação, melhor será a qualidade, visto que tal corresponderá à gravidade média dos problemas detetados em cada segmento que compõe o texto traduzido.

resultados (com e sem “M/M”) seguem a mesma linha do que tinha sido registado no número de ocorrências.

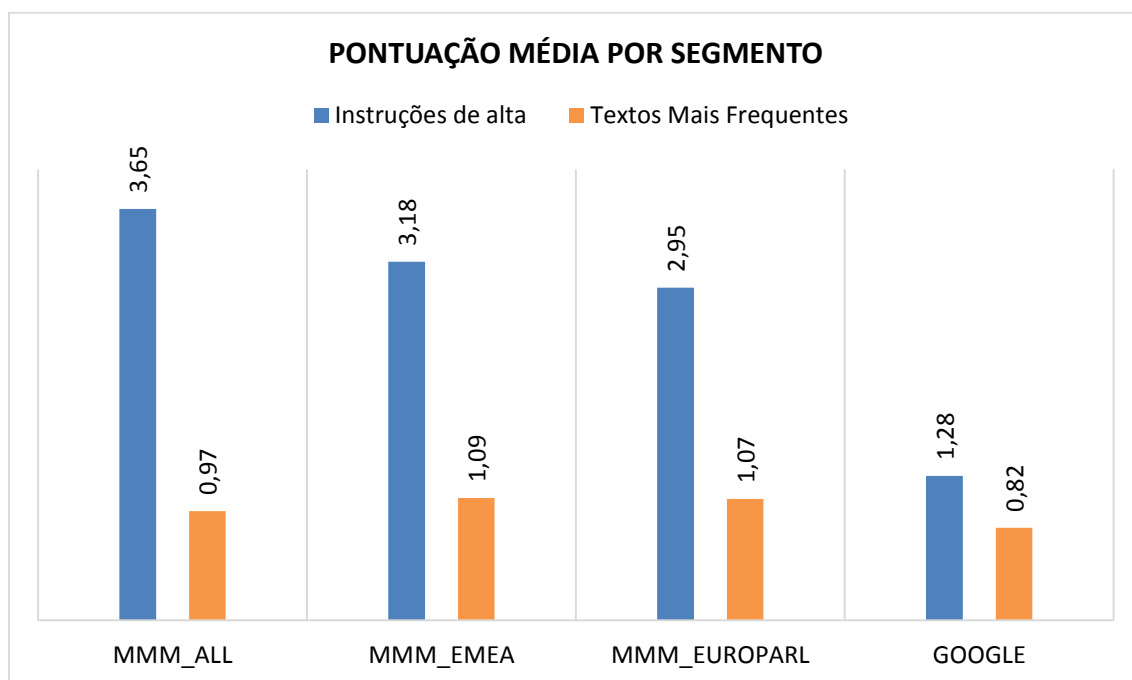


Figura 44 – Pontuação média por segmento nos textos traduzidos pelo MMM e Google Tradutor.

No que diz respeito à tradução do IA, a pontuação acompanhou novamente a dimensão do *corpus* de treino, ou seja, quanto maior foi o *corpus* melhor foi a pontuação: ALL – 3,65, EMEA – 3,18, Europarl – 2,95 e Google – 1,28. Quanto à tradução do TMF, tal como aconteceu com o número de ocorrências, também neste caso a tradução com melhores resultados foi a do Google (0,82), seguida da do ALL (0,97), Europarl (1,07) e EMEA (1,09). Também nesta métrica, a diferença de pontuação entre a tradução do IA e TMF foi mais evidente no caso dos motores MMM (ALL – 36,5/0,97, EMEA – 3,18/1,09 e Europarl – 2,95/1,07) do que no do Google (1,28/0,82).

Tal como aconteceu em relação ao número de ocorrências por texto, também na média de pontuação por segmento, quando foram excluídos os erros de “M/M”, os resultados sofreram algumas alterações relevantes, conforme é possível verificar na Figura 46.

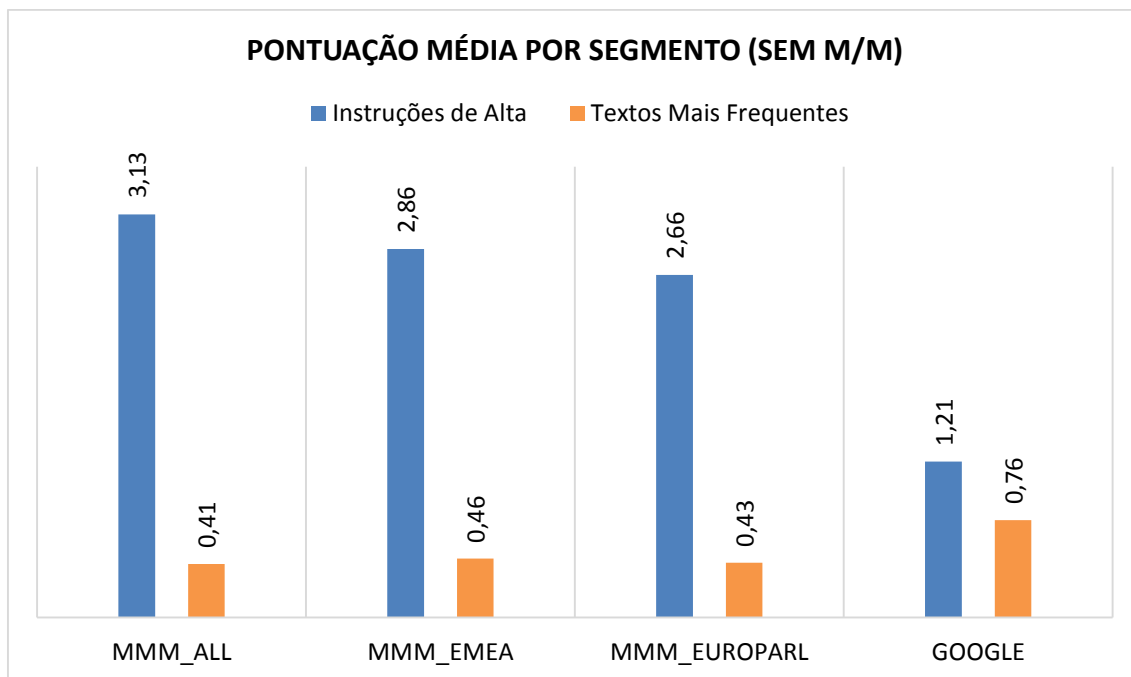


Figura 45 – Pontuação média por segmento (sem contabilizar “M/M”) nos textos traduzidos pelo MMM e Google Tradutor.

Se em termos das traduções do IA a tendência se manteve, ou seja, os sistemas com *corpora* maiores apresentaram melhor pontuação (Google – 1,21, Europarl – 2,66, EMEA – 2,86 e ALL – 3,13), o mesmo não aconteceu com as traduções do TMF. Neste caso (tal como aconteceu no número de ocorrências totais), ao excluir as “M/M”, as traduções com melhor pontuação passaram a ser as realizadas pelo MMM (ALL – 0,41, Europarl – 0,43 e EMEA – 0,46) e o Google passou a ser o sistema com pior pontuação (0,76). Mais uma vez, os resultados entre todos os motores MMM foram muito semelhantes (embora com o ALL a obter o melhor resultado) e revelou-se uma diferença acentuada entre resultados da tradução do IA e TMF (ALL – 3,13/0,41, EMEA – 2,86/0,46, Europarl – 2,66/0,43), o que não aconteceu novamente, pelo menos de forma tão vincada, com o Google (1,21/0,76).

Em termos da distribuição dos níveis de gravidade dos problemas detetados, é possível verificar na Figura 47 que a maioria das ocorrências registadas foram de nível menos grave, ou seja, de nível 1, tanto no caso da tradução do IA (ALL – 49%, EMEA – 51%, Europarl – 47% e Google – 51%) como do TMF (ALL – 85%, EMEA – 86%, Europarl – 87% e Google – 59%).

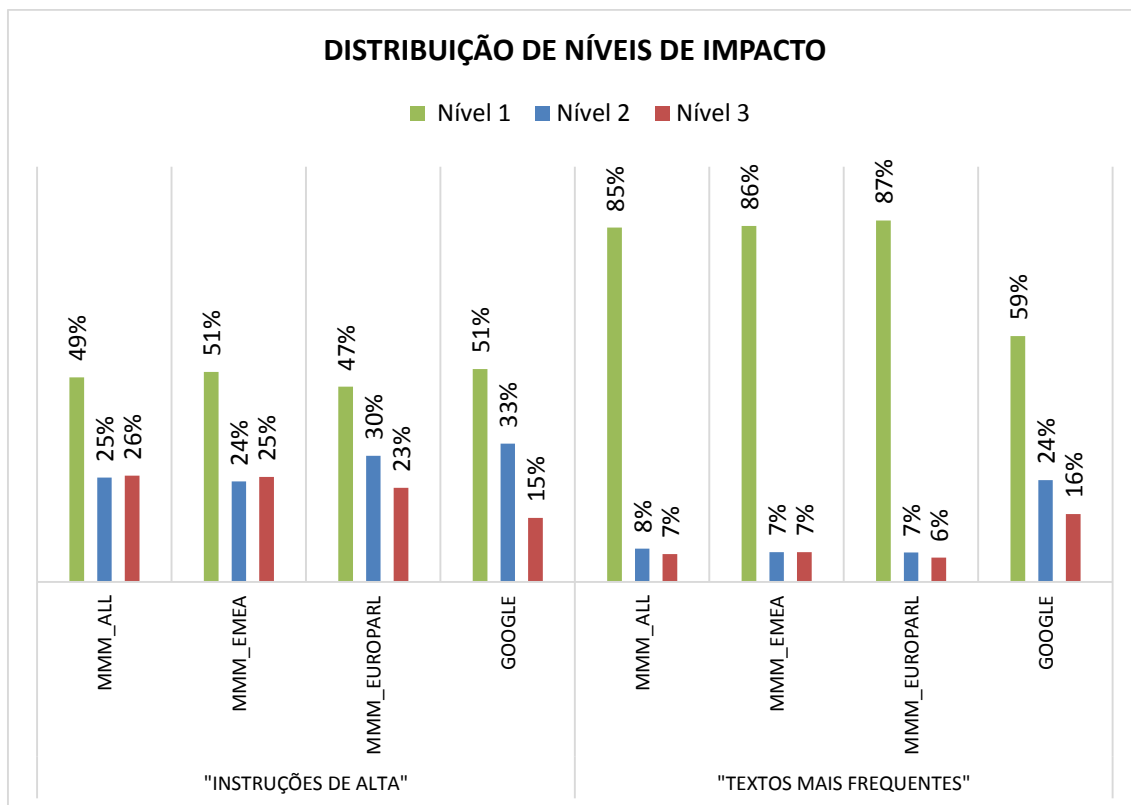


Figura 46 – Distribuição de níveis de impacto por texto traduzido pelo MMM e Google.

Com a exceção do Europarl e do Google, em que há uma diferença que pode ser considerada relevante entre as ocorrências de nível 2 e 3 (Europarl – 30% / 23% e Google – 33% / 15%) na tradução do IA, e apenas do Google (24% / 16%), no caso da tradução do TMF, em todos os outros casos as ocorrências de nível moderado (nível 2) e grave (nível 3) são praticamente equivalentes (IA: ALL – 25% / 26%, EMEA – 24% / 25%; TMF: ALL – 8% / 7%, EMEA – 7% / 7%, Europarl – 7% / 6%). É de notar que, no geral, a percentagem de problemas de nível 2 aumentou com a dimensão do *corpus* (ALL – 25%, EMEA – 24%, Europarl – 30% e Google – 33%), enquanto a percentagem de problemas de nível 3 diminuiu (ALL – 26%, EMEA – 25%, Europarl – 23% e Google – 15%) nas traduções do IA. Quanto à percentagem de ocorrências de nível 1, estas mantiveram-se relativamente semelhantes em todas as traduções. Nas traduções do TMF, a tendência é contrária, ou seja, tanto no caso de ocorrências de nível 2 como 3, estas mantiveram-se estáveis nas traduções do MMM e aumentaram nas traduções do Google. O contrário sucedeu com as ocorrências de nível 1, ou seja, a percentagem de ocorrências deste nível diminuiu na passagem do MMM para Google e mantiveram-se estáveis no MMM.

Quando a categoria “M/M” foi excluída, notaram-se alterações nuns casos e manutenção da tendência, noutros (Figura 48).

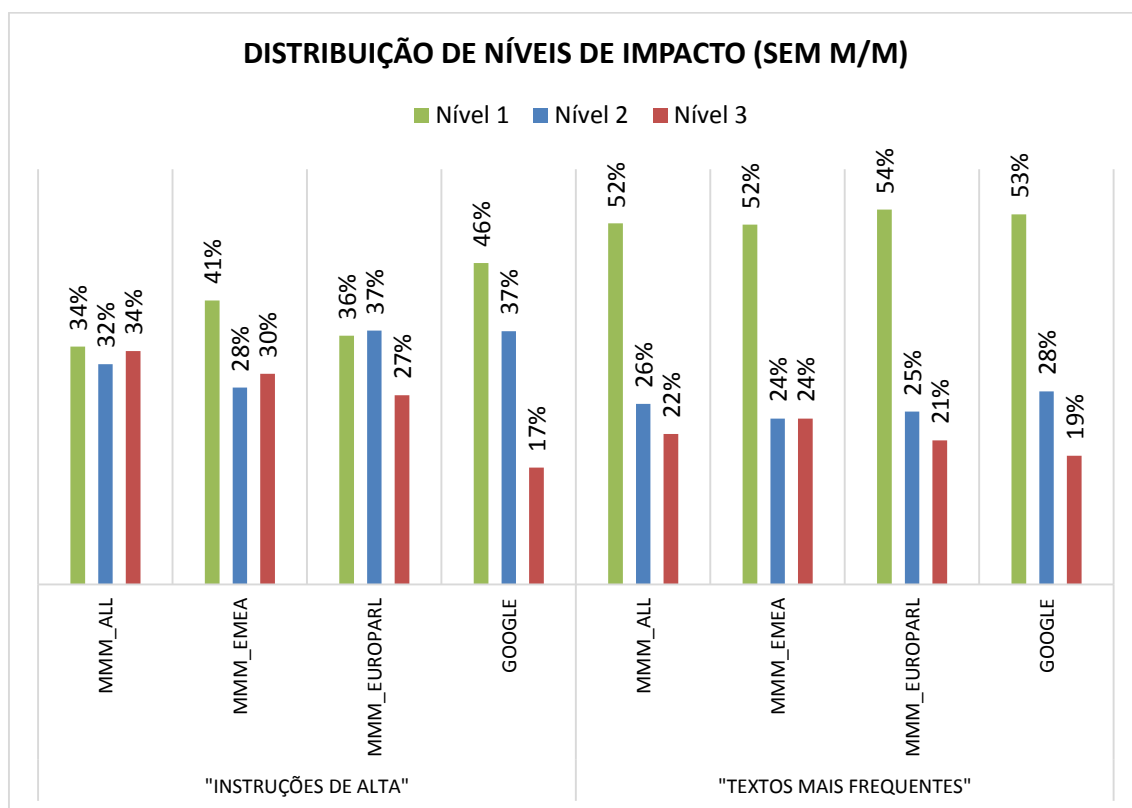


Figura 47 – Distribuição de níveis de impacto por texto traduzido pelo MMM e Google (sem contabilizar “M/M”).

Também neste caso, a grande maioria das ocorrências registadas pelas traduções do MMM e Google foi de gravidade mais baixa (nível 1), especialmente na tradução do TMF. As exceções foram as traduções de IA realizadas pelo ALL (34%) e Europarl (37%). No caso do ALL, existiu uma distribuição quase perfeita entre os três níveis de gravidade (34% – 32% – 34%), nesta tradução. No caso da tradução do IA do Europarl, a percentagem de ocorrências de nível 1 (36%), embora inferior, foi quase semelhante à de nível 2 (37%). Em quase todos os casos, a percentagem de erros de nível 2 foi superior à percentagem de erros de nível 3. As únicas exceções foram as traduções do IA realizadas pelo ALL e pelo EMEA (ALL – 32% / 34% e EMEA – 28% / 30%, respetivamente) e a tradução do TMF realizada pelo EMEA (24%/24%).

No que diz respeito às categorias dos problemas detetados (Tabela 3), aquelas que registaram maior número de ocorrências em ambos os textos foram “Maiúsculas/minúsculas” (295), “Gramática” (219), “Tradução” (179) e “Omissão/adição” (118). As restantes categorias, apresentam um número

consideravelmente menor de ocorrências: “Coerência” (42), “Pontuação” (40) e “Terminologia” (35).

Tabela 3 – Ocorrências totais por categorias de problemas.

		GRAM.	TERM.	PONT.	OM/AD	TRAD.	MAI/MIN	COER.	TOTAL P/ SISTEMA
INSTRUÇÕES DE ALTA	MMM_ALL	56	7	11	29	48	50	6	207
	MMM_EMEA	45	6	9	40	36	31	15	182
	MMM_EUROPARL	47	7	10	23	41	29	11	168
	GOOGLE	26	3	4	7	29	7	2	78
	SUBTOTAL	174	23	34	99	154	117	34	635
TEXTOS MAIS FREQUENTES	MMM_ALL	9	2	1	3	6	52	2	75
	MMM_EMEA	9	2	3	3	7	59	1	84
	MMM_EUROPARL	10	1	1	2	8	61	2	85
	GOOGLE	17	7	1	11	4	6	3	49
	SUBTOTAL	45	12	6	19	25	178	8	293
TOTAL		219	35	40	118	179	295	42	

Na tradução de IA, as categorias com maior número de registos foram “Gramática” (174), “Tradução” (154), “Maiúsculas/minúsculas” (117) e “Omissão/adicação” (99). Na tradução de TMF a maioria dos problemas registados foram relativos às seguintes categorias: “Maiúsculas/minúsculas” (178), “Gramática” (45), “Tradução” (25) e “Omissão/adicação” (19). As restantes categorias, apresentaram, tanto na tradução de um como de outros textos, bastante menos ocorrências. De relevar apenas que se na tradução de IA as questões de “Terminologia” foram as que menos problemas registaram (23), o mesmo não aconteceu na tradução de TMF (12).

5.2. Resultados da avaliação automática

A avaliação das traduções realizada através do sistema de avaliação automática do MMM revelou os resultados que são apresentados na Figura 49, onde é apresentada a pontuação BLEU global de cada texto.

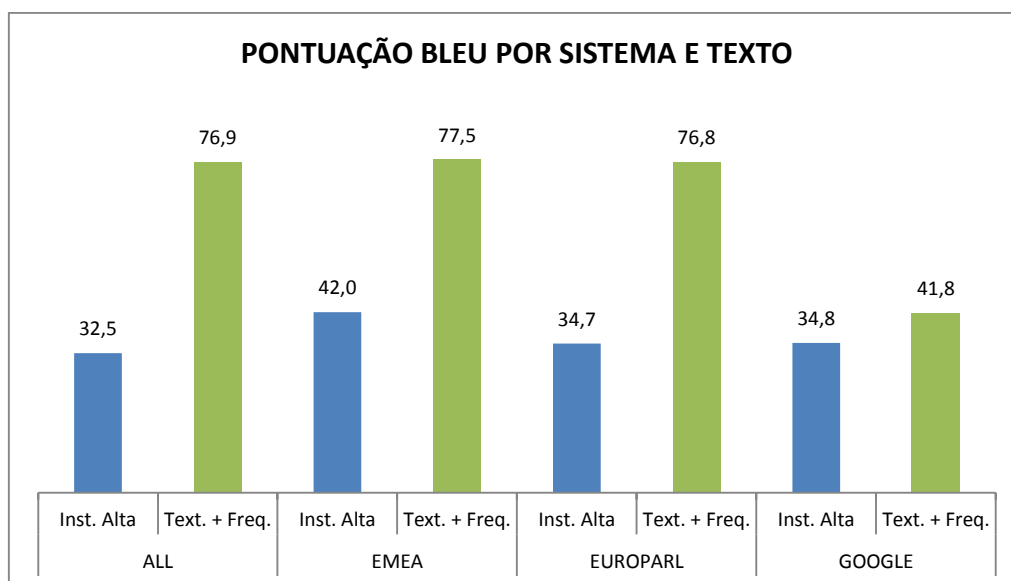


Figura 48 – Pontuação BLEU por sistema e por texto.

No que diz respeito à tradução das “Instruções de Alta” (IA), aquela que apresentou melhor pontuação BLEU foi a tradução realizada pelo MMM com o *corpus* EMEA (42,0). A segunda melhor pontuação pertenceu ao Google (34,8), com uma ligeira diferença em relação à do Europarl (34,7). A pior tradução, de acordo com a pontuação BLEU, pertenceu ao ALL (32,5).

Quanto às traduções dos “Textos Mais Frequentes” (TMF), a melhor pontuação pertenceu novamente ao EMEA (77,5), seguido do ALL (76,9) e do Europarl (76,8). A tradução com pior resultado BLEU foi a do Google (41,8).

Se na tradução do IA, tanto as traduções do MMM como do Google apresentaram resultados muito semelhantes (a diferença entre o Google e a tradução MMM com pontuação mais próxima – Europarl – foi de 0,1 ponto), o mesmo não se passou na tradução de TMF. Neste último caso, a diferença entre a pontuação obtida pela tradução do Google e a tradução do MMM mais próxima (Europarl) foi de 35,1 pontos. Por outro lado, as traduções do MMM apresentaram resultados muito semelhantes entre si, principalmente ao nível da tradução do TMF: a diferença entre a melhor e a pior tradução do TMF foi de apenas 0,6 e, em relação ao IA, foi de 9,5.

De notar ainda a diferença entre as pontuações BLEU do IA e TMF: as traduções do TMF apresentaram resultados bastante superiores em relação às traduções do IA, em especial no que diz respeito às traduções do MMM. A menor diferença registada entre traduções MMM de IA (EMEA – 42) e TMF (EMEA – 77,5) foi de 35,5. Nas traduções do Google, por seu lado, embora a tradução do TMF também apresentasse melhor

resultado do que a do IA, a diferença entre estas não foi tão marcada: a diferença entre a tradução do Google do IA (34,8) e do TMF (41,8) foi de apenas 7 pontos.

6. Discussão de Resultados

Analisando a prestação de MMM e Google, é possível verificar, tanto através da avaliação automática como da avaliação manual, que, na generalidade, ambos os sistemas tiveram mais dificuldade em traduzir o documento IA do que o TMF. Isto expressa-se não tanto pela quantidade de ocorrências⁸⁴, mas principalmente pela qualidade do resultado final.

Em termos de relação da quantidade de ocorrências com a dimensão do *corpus* de treino de cada sistema, na tradução do IA (tanto com como sem “M/M”), o número de ocorrências foi inversamente proporcional à dimensão do *corpus* de treino, ou seja, quanto maior foi o *corpus*, menor foi o número de ocorrências registadas e vice-versa (com “M/M”: ALL – 207, EMEA – 182, Europarl – 168 e Google – 78; sem “M/M”: ALL – 157, EMEA – 151, Europarl – 139 e Google – 71) (cf. Figuras 43 e 44). Na tradução do TMF, quando considerada a categoria “M/M”, continuou a ser o sistema com o *corpus* de treino maior (Google) aquele que obteve o melhor resultado (49 ocorrências) (cf. Figura 44). Contudo, entre as traduções produzidas pelos diferentes motores MMM, registou-se uma relação direta e proporcional entre a dimensão do *corpus* e a quantidade de ocorrências, ou seja, quanto maior era o *corpus* maior foi o número de número de ocorrências registadas (ALL – 75, EMEA – 84 e Europarl – 85) (cf. Figura 43). Quando foi excluída a categoria “M/M”, esta tendência tornou-se ainda mais evidente e alargou-se ao Google, passando a ser a tradução deste sistema a deter o maior número de ocorrências (ALL – 23, EMEA – 25, Europarl – 24, Google – 43) (cf. Figura 44).

O mesmo aconteceu na pontuação média por segmento (Figura 45 e 46), a medida utilizada para obter uma noção imediata e muito genérica acerca da qualidade da tradução (quanto menor a pontuação, melhor a qualidade, visto que tal corresponderá à gravidade média dos problemas detetados em cada segmento). Também este indicador apresentou uma tendência semelhante à registada na contagem global de ocorrências na tradução de IA: se na tradução do IA, tanto com como sem “M/M”, a pontuação diminuiu de acordo com a dimensão dos *corpora* (com “M/M”: ALL – 3,65, EMEA – 3,18, Europarl – 2,95 e Google – 1,28; sem “M/M”: ALL – 3,13, EMEA – 2,86,

⁸⁴ Apesar de as traduções do IA apresentarem o dobro das ocorrências registadas do TMF (635 e 293, respetivamente), esses resultados são proporcionais com a dimensão de cada texto (1.1142 palavras o IA e 547 o TMF) – a percentagem de ocorrências é 56% nas traduções do IA e de 54% no caso do TMF.

Europarl – 2,66 e Google – 1,21) (cf. Figura 45 e 46), na tradução do TMF os resultados oscilaram e apontaram num sentido inverso. Quando contabilizada a categoria “M/M”, a melhor tradução continuou a ser aquela que foi realizada pelo sistema com o maior *corpus* de treino – ou seja, o Google (0,82) –, embora no âmbito do MMM se notasse uma tendência inversa, ou seja, de os melhores resultados seguirem numa direção oposta à da dimensão do *corpus* (ALL – 0,97, EMEA – 1,09 e Europarl – 1,07) (cf. Figura 45 e 46). Ao excluir a categoria “M/M” do cálculo da pontuação média por segmento notou-se de forma mais evidente esta última tendência, ou seja, as traduções com melhor pontuação foram as realizadas com base em *corpora* mais pequenos: ALL – 0,41, EMEA – 0,46, Europarl – 0,43 e Google – 0,76 (cf. Figura 46). Aliás, tanto na contabilização de ocorrências como na pontuação média por segmento, a diferença de resultados entre os diferentes motores MMM na tradução do TMF foi tão pequena que se poderá considerar que esta não é significativa e que os 3 motores apresentaram resultados semelhantes. Este facto parece indicar que o aumento de dados ao *corpus* base – ALL – não produziu efeitos relevantes na tradução do TMF, independentemente da quantidade ou pertinência dos dados acrescentados.

O facto de as traduções de ambos os textos apresentarem resultados tão díspares pode estar relacionado com dois fatores: 1) o facto de uma parte considerável do texto TMF possuir vários segmentos com um grau elevado de semelhança em relação aos segmentos que compõem o *corpus* ALL, o que não acontece com o texto IA⁸⁵; e 2) o tipo de texto e a sua função na aplicação ALERT®. Os textos que compõem o TMF são eminentemente técnicos, com muita terminologia e fraseologia típica da área clínica. Além disso, trata-se de um conjunto de textos que não estão necessariamente relacionados uns com os outros por serem utilizados em diferentes áreas do ALERT® e por diferentes profissionais (médicos, enfermeiros, administrativos, etc.). Isto traduziu-se em segmentos incompletos para a inserção de informação por parte do utilizador, títulos de campos, e segmentos em forma de comentários breves com frases que não obedecem à estrutura típica do português (Sujeito-Verbo-Objeto ou SVO). O texto IA, por seu lado, constitui um conjunto de segmentos relacionados entre si e com coesão e coerência internas. Sendo um texto utilizado para transmitir a pacientes, possui menos terminologia e fraseologia clínica, em comparação com o TMF. Além disso, os segmentos correspondem, na sua grande maioria, a frases completas com a estrutura

⁸⁵ Cf Anexo xi, pp. 126.

SVO. Dado tudo isto, é compreensível que sistemas menos orientados a textos como o TMF, por serem muito específicos, possuam maiores dificuldades na tradução destes.

Em termos da distribuição da gravidade das ocorrências registadas na tradução de cada texto, tanto com como sem “M/M”, verifica-se que grande parte das ocorrências (principalmente nas traduções do MMM de TMF com “M/M”) foram de nível 1 e a parte restante encontra-se repartida, de forma muito equilibrada, entre os níveis 2 e 3 (Figuras 47 e 48). Isto parece revelar que não há uma relação direta e significativa entre a dimensão do *corpus* e a distribuição da gravidade dos problemas detetados nas traduções de ambos os documentos.

No que diz respeito à distribuição das ocorrências pelas categorias de erros (Tabela 3), é possível verificar que, no geral, as principais áreas problemáticas na tradução de ambos os textos foram as seguintes: “Maiúsculas/Minúsculas” (295), “Gramática” (219), “Tradução” (179) e “Omissão/Adição” (118). A problemática das “M/M” foi especialmente notada nas traduções do MMM e com maior incidência nas traduções do TMF. Apesar de esta categoria não ter implicações gramaticais na tradução, nem constituir um elemento de distorção da mensagem (como acontece, por exemplo, com erros de pontuação), portanto, por não influenciar a qualidade da tradução ou a gramaticalidade de um segmento, é um critério de avaliação da qualidade de qualquer tradução. A incapacidade de um sistema de TA devolver um resultado com a devida integração de maiúsculas e minúsculas, revela um de dois fatores: a incapacidade do seu componente *Recaser*⁸⁶ em criar um modelo fiável ou a necessidade de alterar parâmetros num dos *scripts* do sistema. De qualquer forma, esta categoria não define a qualidade intrínseca da tradução, mas apenas a qualidade do resultado final que é disponibilizado a um utilizador final. Daqui pode-se, porventura, inferir que o MMM parece ter sido concebido mais para criar versões iniciais de traduções para serem pós-editadas, do que traduções finais prontas a serem disponibilizadas a utilizadores também eles finais (como acontece com o Google, em que há muitas menos ocorrências de “M/M” e “Pontuação”, uma vez que este sistema parece estar mais bem preparado e orientado para manter a formatação do TP no TC).

⁸⁶ Para traduzir um texto, um sistema de TAE normalmente reduz todas as palavras a minúsculas. O *Recaser* é o componente de um sistema de TAE que repõe as maiúsculas e minúsculas num produto final do sistema, criando um modelo para o efeito que é aprendido através de um *corpus* de treino. Para mais informações acerca deste componente consultar <http://www.statmt.org/moses/?n=Moses.SupportTools#ntoc10> (último acesso em setembro de 2014).

Além das “M/M”, a “Gramática” foi outra das áreas em que ambos os sistemas sentiram maiores dificuldades, em ambos os textos. Em termos de “Gramática”, as traduções do IA revelaram que à medida que *corpus* crescia, em geral, o número de ocorrências diminuía (*vide* Tabela 3). No caso das traduções do TMF, ocorreu o contrário, ou seja, à medida que o *corpus* crescia, o número de ocorrências aumentava (*vide* Tabela 3). Esta variação entre as traduções dos dois textos pode estar diretamente ligada com o tipo de texto, tal como referido anteriormente, ou seja, para textos mais genéricos é útil um *corpus* maior, para textos mais técnicos e orientados especificamente a áreas do ALERT® é útil um *corpus* mais orientado e não tanto um *corpus* extenso.

Apesar de o MMM apresentar mais problemas de “Tradução” do que o Google em ambos os textos, as ocorrências entre ambos os sistemas nunca variaram tanto quanto aconteceu na generalidade das restantes categorias (*vide* Tabela 3). A maior diferença entre os sistemas ocorreu nas categorias “M/M” e “Omissão/adição”. Se o caso das “M/M” já foi analisado, no caso da “Omissão/adição” verificou-se que nas traduções do IA foram os motores MMM a registar mais ocorrências nesta categoria do que o Google. Contudo, no caso das traduções do TMF observou-se o contrário. Esta discrepância de resultados entre traduções de ambos os textos pode indicar questões de n-gramas. Se nas traduções do IA o MMM pareceu ter mais dificuldades, possivelmente por falta de dados que lhe permitissem detetar n-gramas da mesma ordem, o mesmo aconteceu ao Google na tradução do TMF. Mesmo sendo ambos os textos bastante técnicos, o TMF tem mais especificidades, próprias dos textos para o ALERT®, pelo que parece ter sido essa uma das razões pelas quais o Google não foi capaz de manter a mesma informação do TP para o TC.

Outras das categorias em que se registou uma discrepância relevante de ocorrências entre o Google e o MMM foi ao nível da “Terminologia” (*vide* Tabela 3). Enquanto o Google não teve muitos problemas na tradução do IA, por ser um texto composto por terminologia mais corrente, destinado a pacientes, no caso do TMF, por serem textos para profissionais de saúde e a terminologia ser mais densa, o Google apresentou mais dificuldades. Este facto pode dever-se ao Google não ser um sistema orientado a um domínio específico do conhecimento, ao contrário do MMM, que utilizou *corpora* que tiveram na sua base traduções semelhantes àquelas que são apresentadas no TMF.

Todos estes dados recolhidos a partir da avaliação manual parecem apontar para a necessidade de equilíbrio entre quantidade e “qualidade” (*i.e.* pertinência dos dados) dos *corpora* na tradução dos textos IA e TMF para o ALERT®. A quantidade não parece

acrescentar qualidade quando o conteúdo não é orientado ao domínio do texto e ao tipo de texto, assim como a pertinência do *corpus* não parece ser suficiente para dar conta de textos que, mesmo sendo do mesmo domínio, possuam conteúdo desconhecido do *corpus*.

A interpretação que se pode fazer dos resultados da avaliação automática também não está longe da conclusão retirada da avaliação manual. Neste caso, a métrica BLEU elege o *corpus* EMEA como o melhor na tradução tanto do IA como do TMF. Isto pode querer indicar que é necessário um misto de “qualidade” e quantidade no *corpus* de treino para se produzirem melhores resultados, em vez de ter apenas um *corpus* com “qualidade” e pouca quantidade e vice-versa. Neste caso, o facto de o EMEA obter os melhores resultados pode estar relacionado com o facto de este ser o *corpus* com mais quantidade de conteúdos clínicos. Com o aumento de dados com o Europarl, os resultados decresceram, assim como com a redução do *corpus* a apenas ao ALL, (embora a diferença entre ALL, EMEA e Europarl na tradução do TMF, seja demasiado pequena para se considerar relevante – ou seja, nesse caso, o aumento da quantidade parece não ter resultado em ganhos evidentes em termos de qualidade). Por outro lado, o Google, sendo (de longe) o sistema com o maior *corpus* de treino, aproxima-se mais dos resultados do Europarl, que é também aquele que possui maior quantidade de dados fora do domínio da tradução, do que dos restantes motores MMM.

Todos os dados recolhidos, tanto da avaliação manual como automática, parecem indicar que, no caso em estudo, o MMM traduz melhor os textos que são mais semelhantes ao seu *corpus* de treino e que o acréscimo de dados, quando não é da mesma área de conhecimento, nem ainda do mesmo tipo de textos, não representa um acréscimo de qualidade na tradução. Mesmo os problemas gramaticais foram mais bem ultrapassados pelo MMM quando traduziu textos de domínio mais restrito (*in-domain*) do que o Google – isto parece indicar que a “qualidade” (*i.e.* a pertinência dos dados) do *corpus* tem influência mesmo no desempenho de questões gramaticais. Por outro lado, o Google foi capaz de obter melhor desempenho num texto que, mesmo sendo do mesmo domínio de conhecimento da base dos *corpora* MMM, apresentava menos semelhanças com o *corpus* base do MMM. Como tal, daqui se conclui que “nem a gordura é formosura” nem a “magreza é beleza” na TAE. Tal como em quase tudo na vida, também nos sistemas TAE “a virtude encontra-se no meio”, ou seja, o ideal será uma combinação de quantidade e “qualidade” do *corpus* de treino.

7. Considerações Finais

Com este estudo de caso procurou-se testar aquilo que se considera “sabedoria convencional”⁸⁷ em relação a sistemas de TAE, ou seja, a crença de que “mais dados é melhor; quanto maior for um *corpus* de treino, mais exato será o modelo”⁸⁸ (Axelrod, Xiaodong, & Jianfeng, 2011: 1). Mais propriamente, procurou-se perceber se na tradução de textos de um domínio específico é mais importante a quantidade ou a pertinência dos dados do *corpus* bilingue usado no treino do sistema de TAE.

Começou-se por fazer um enquadramento da TA no mundo atual, em que a tecnologia é vista por autores como Nicholas Ostler (2010) como forma de superar as limitações humanas, nomeadamente as barreiras linguísticas. Atualmente a tecnologia está presente e tem um impacto evidente e crescente na vida quotidiana de cada um. A área das ciências da vida é uma das que mais tem beneficiado do progresso tecnológico. Muito do progresso atual é baseado na quantidade de informação que é criada, processada e partilhada. Nesse âmbito, surgem aplicações informáticas, como o ALERT®, que se dedicam à gestão da informação em ambientes clínicos. É neste contexto, da era da informação, que ferramentas de TA adquirem um grande valor para apoiar a partilha rápida e universal da informação através da tradução.

De forma a se compreender o ponto atual da investigação na TA e as expectativas que recaem sobre a mesma, é importante conhecer o percurso desta. Como tal, procedeu-se a um resumo histórico do caminho traçado pela TA: desde as grandes expectativas criadas aquando do estabelecimento desta linha de investigação na área computacional durante a década de 50 do século passado; passando pelo ceticismo da década seguinte em resultado da incapacidade da investigação em TA atingir os objetivos ambiciosos a que se tinha proposto; pelas primeiras implementações de sistemas de TA na década de 80; pelo início da comercialização de ferramentas de TA a empresas de diversos setores; até ao grande avanço e expansão desde o final do século passado graças à disponibilização de grandes quantidades de texto em formato digital através da Internet. No âmbito da investigação da TA apresentaram-se ainda as diferentes abordagens mais comuns: sistemas baseados em regras, sistemas baseados em estatística e sistemas híbridos, que aplicam ambas as abordagens anteriores. Destacou-se ainda a importância da avaliação

⁸⁷ Tradução nossa.

⁸⁸ Tradução nossa.

da TA, tanto manual como automática, na orientação e avanço da investigação. A avaliação de TA adquiriu uma importância tal ao ponto de se ter tornado uma linha de investigação independente dentro da investigação em TA.

A seleção de sistemas de TAE para este projeto teve a ver com o facto de esta ser atualmente o paradigma dominante na investigação de TA (de acordo com Specia, 2013: 74). Dessa forma, ao longo de um capítulo inteiro, procedeu-se a uma descrição geral do funcionamento dos sistemas de TAE, da teoria que subjaz aos seus componentes principais (modelo de tradução, decodificador, modelo de língua) e ainda de algumas das formas de aquisição de conhecimento linguístico. A descrição de cada um dos componentes e das formas de aquisição de conhecimento linguístico processou-se de forma muito genérica, tendo em consideração não só o facto de se tratar de uma tese de mestrado de Linguística (e não de ciências informáticas), como ainda tendo em conta as nossas próprias limitações no que diz respeito a conhecimentos estatísticos e informáticos. Por esse motivo, procurámos evitar, sempre que possível, transcrever fórmulas matemáticas. Em vez disso, optámos por explicar os princípios teóricos e consequências práticas por trás dessas fórmulas aplicadas ao sistema de TAE.

Feita a contextualização teórica, procedeu-se à descrição dos materiais e métodos usados no estudo de caso. Para este projeto foram selecionados dois tipos de sistema de TAE: o Moses for Mere Mortals, um *software* livre e de código aberto, disponível para descarregamento gratuito na Internet e com capacidade para ser treinado com *corpora* próprios do utilizador; e o Google Tradutor, uma aplicação de utilização gratuita (dentro de determinados limites) em linha, que utiliza grandes *corpora* para o seu treino, mas que não permite o treino com *corpora* próprios. De entre os vários obstáculos que foi necessário superar para realizar este projeto, destacam-se as dificuldades com a instalação e interação com o MMM. Apesar de ser uma versão simplificada do Moses – que é uma aplicação essencialmente construída por e para engenheiros informáticos/linguistas computacionais –, o conjunto de conhecimentos necessários para operar esta ferramenta ultrapassa ainda assim largamente os meros conhecimentos informáticos “na ótica do utilizador”, que são requisito da grande maioria das aplicações comerciais. Para operar o MMM foi necessário adquirir conhecimentos no sistema de código aberto Linux (para a instalação e utilização do mesmo); na sintaxe da linha de comandos Linux; conhecimentos de compatibilidades e incompatibilidades entre componentes informáticos; e conhecimentos mais profundos sobre o funcionamento prático de um sistema de TAE e seus componentes. Não fosse o apoio de várias outras

peessoas com mais ou menos conhecimentos acerca do MMM, a persistência através de sucessivas tentativas-erro, e teria sido, se não impossível, pelo menos extremamente demorada a aprendizagem do funcionamento desta aplicação, uma vez que não é uma aplicação de todo simples, pois não tem interface gráfica que permita a interação com o utilizador. Todas as ações são realizadas através de *scripts* anotados, mas que de pouco servem para quem não domine bem os conceitos e componentes básicos dos sistemas TAE.

Foram criados três motores MMM e treinados com três tipos de *corpora*. O primeiro *corpus* (ALL), que serviu de base aos restantes, foi composto por memórias de tradução da ALERT (ou seja, tratou-se de um *corpus* orientado a uma área específica dentro do domínio da saúde); o segundo, além de conter o *corpus* base, incluiu ainda o *corpus* da *European Medicines Agency* (EMA) (ou seja, um *corpus* que, mesmo não sendo tão orientado ao tipo de texto a traduzir, faz parte do domínio da saúde); e o terceiro, além de conter os outros dois *corpora*, incluiu ainda o *corpus* Europarl, que consiste em transcrições das sessões do Parlamento Europeu (ou seja, um *corpus* de um domínio geral, não orientado especificamente à área da saúde). Os textos a traduzir foram textos que são utilizados nos produtos clínicos ALERT®, mas que nunca tinham sido traduzidos antes. Nesse âmbito foram selecionados dois textos: “Instruções de alta” (IA), que é composto por indicações de profissionais clínicos para o paciente com informação de saúde do problema diagnosticado ao paciente; e “Textos Mais Frequentes” (TMF), que é um conjunto de textos normalmente utilizados por profissionais de saúde para registar informação clínica sobre pacientes e exames e orientados à partilha entre profissionais. Embora sejam ambos do domínio da saúde, o primeiro utiliza uma linguagem mais acessível ao público em geral, por ser direcionado a não-especialistas, enquanto o segundo, sendo igualmente do domínio da saúde, é muito específico da aplicação ALERT®, por se destinar a áreas específicas da aplicação informática que apenas são utilizadas por profissionais de saúde. Desse modo, poder-se-ia considerar que o TMF pertence a um grupo de conteúdo clínico mais restrito. Ambos os textos foram analisados antes da tradução com uma ferramenta de tradução apoiada por computador (SDL Trados Studio 2014) e verificou-se que o IA praticamente não apresentava segmentos semelhantes aos dados presentes nas memórias de tradução da ALERT (que serviram para criar o *corpus* ALL). O TMF, por seu lado, apesar de nunca ter sido traduzido anteriormente, apresentava bastantes semelhanças com o conteúdo presente na memória de tradução ALERT. A avaliação das traduções produzidas pelos

dois sistemas foi realizada de duas formas: automática e manual. Para a avaliação automática, utilizou-se a ferramenta de avaliação automática do MMM, que emprega as métricas BLEU e NIST. Optou-se por apenas utilizar a métrica BLEU, não só por esta ser a mais comum neste tipo de avaliações, mas ainda porque se considerou que a métrica NIST pouco iria acrescentar, uma vez que esta constitui uma variação da métrica BLEU. Para a avaliação manual optou-se por criar um sistema próprio, visto que, após investigação às ferramentas e sistemas de avaliação manual de traduções, se concluiu que nenhuma se adaptava às necessidades do projeto. O sistema de avaliação manual foi composto por dois grupos de classificação: uma classificação dos problemas por categorias (problemas gramaticais, terminológicos, de pontuação, omissão/adição de informação, tradução, má utilização de maiúsculas e/ou minúsculas, e coerência) e outra de acordo com o nível de gravidade (nível 1, 2 e 3 ou ligeiro, moderado e grave). Apesar de termos procurado criar uma métrica de avaliação manual que se adaptasse às necessidades do projeto e de tentar que os critérios utilizados para essa avaliação fossem o mais objetivos possível (nomeadamente através da definição de cada uma das categorias de problemas e dos níveis de gravidade), temos a consciência de que os resultados desta avaliação dependem sempre, em certa medida, do juízo individual de cada avaliador. Se casos houve em que determinado erro poderia pertencer a mais do que uma categoria de erro ou merecer outro nível de gravidade, procurou-se, no entanto, adotar uma posição coerente, atribuindo a erros semelhantes a mesma categoria e nível de gravidade em todos os casos e em todas as traduções. A utilização de uma métrica manual e outra automática teve como objetivo obter perspetivas diferentes acerca da qualidade das traduções. Se com a avaliação é feita uma análise geral do texto traduzido em comparação com outra tradução de referência, a avaliação manual permite uma análise mais detalhada a cada segmento do texto traduzido. Além disso, a utilização de duas métricas, uma mais geral e outra mais pormenorizada, pode permitir analisar até que ponto os resultados de ambas se encontram em sintonia.

Na avaliação dos resultados, ressaltou a quantidade de ocorrências de problemas relacionados com maiúsculas e minúsculas, de tal forma, que as ocorrências desta categoria de erro acabaram por influenciar de forma desviante os resultados gerais das traduções do MMM, nomeadamente do TMF. Nesse sentido, os resultados foram analisados com e sem o registo de ocorrências com maiúsculas e minúsculas (“M/M”), visto que este tipo de problema não influencia o juízo de gramaticalidade de uma frase nas línguas em questão (português e inglês), nem contribui para a distorção da

informação na passagem do TP para o TC. Neste contexto, foi possível verificar que na tradução do IA, texto do domínio clínico, tiveram melhores resultados os sistemas que utilizaram *corpora* maiores. Na tradução do TMF, considerado um texto mais específico dentro do domínio da saúde, foi possível detetar uma tendência inversa: se com o registo de “M/M” se verificou que continuou a ser o sistema com o maior *corpus* de treino (Google) a apresentar melhores resultados, por outro lado, na tradução com o MMM verificou-se que foram os *corpora* mais pequenos, mas mais orientados ao domínio, que apresentaram melhores resultados. Ao analisar as mesmas traduções, mas desta vez sem os registos de problemas com “M/M” tornou-se ainda mais evidente esta última tendência, com os *corpora* mais orientados ao domínio (ALL, EMEA e Europarl) a apresentarem melhores resultados do que o Google. Além disso, foi ainda possível verificar, principalmente na tradução do TMF, que a adição de dados ao *corpus* ALL não resultou em ganhos proporcionais em qualidade. Estes resultados apontam para a necessidade de os *corpora* bilingues utilizados para treinar sistemas TAE apresentarem um equilíbrio entre qualidade e “qualidade” (aqui entendida como pertinência dos dados em relação ao domínio do conhecimento do texto a traduzir) de forma a produzirem melhores resultados.

Os resultados da avaliação automática com a métrica BLEU, apesar de aparentemente divergirem dos resultados da avaliação manual (principalmente na tradução do IA), sublinham a preponderância do equilíbrio entre a “qualidade” e a quantidade. Ao considerar as traduções do EMEA, tanto do IA como do TMF, como as melhores de entre os restantes candidatos, está-se a afirmar que o melhor sistema é aquele que possui mais dados do domínio do texto a traduzir e não aquele que possui simplesmente mais dados (como é o caso do motor baseado no Europarl e do Google). Aliás, os resultados do EMEA acabam por ser bastante próximos do ALL, na tradução de TMF, o que também permite reforçar a ideia de que a adição de dados, se estes não forem da área específica de um domínio, não produzem resultados significativamente melhores.

Este projeto pode servir de complemento a estudos como o de Zhang, *et al.* (2014) que se debruçam sobre as diferentes formas de melhorar o desempenho de sistemas TAE na tradução de textos técnicos. No seguimento deste estudo de caso, poderia ser igualmente útil comparar o desempenho de um sistema de TAE, como o MMM, com um sistema híbrido na tradução de textos técnicos de um domínio específico.

Os resultados deste estudo de caso podem ainda servir para orientar a implementação de um sistema de TAE, como o MMM, no Departamento de Línguas da ALERT, com o objetivo de obter ganhos na produtividade do Departamento.

No que diz respeito a aspetos a melhorar, muito haveria a apontar, pois que se, tal como refere o poeta, “o caminho faz-se caminhando”, também na exploração dos meandros da TA e seus sistemas a experiência empírica é fundamental para a compreensão do melhor caminho a seguir para aproveitar as potencialidades destas ferramentas. Nesse âmbito, sentimos que apenas depois de termos realizado esta experiência estaríamos efetivamente prontos para começar a nossa “caminhada”. Olhando para trás conseguimos ver as nossas falhas com mais clareza e apontá-las com o intuito de encarar esta experiência como uma parte importante do caminho – aquela onde se erra – para depois se poder tentar de novo e melhor. Assim, o principal aspeto a melhorar numa experiência futura semelhante seria selecionar um texto como o TMF que não tivesse tantas ocorrências e tão semelhantes em relação às memórias de tradução que foram usadas como base dos corpora MMM. Neste caso, seria importante que ambos os sistemas em comparação partissem de bases semelhantes, ou seja, em que o grau de semelhança com o conteúdo com que o sistema foi treinado fosse o mais baixo possível, de forma que nenhum dos sistemas apresentasse logo à partida uma vantagem em relação ao outro. Além disso, teria sido também importante proceder a um pré-processamento dos conteúdos com os quais o MMM foi treinado de forma a eliminar problemas ortográficos, gramaticais, de coerência terminológica e outros que influenciassem negativamente as traduções produzidas. Alguns dos erros detetados nas traduções do MMM (como, por exemplo, “adequanda”, “for”, caracteres especiais como “¿”) resultam de ocorrências nos *corpora* com os quais foi treinado. Também os textos de partida deveriam ter sido pré-processados para facilitar a tradução pelos sistemas de TAE. Isto incluiria normalizar formatações (por exemplo, do travessão, com ou sem espaço), pontuação (exemplo, “Dentition:.” e “Dentition:”) ou eliminar caracteres especiais (como “¿”).

Noutro âmbito, poderia ter sido útil selecionar apenas os conteúdos clínicos específicos das memórias de tradução que interessasse compilar num único *corpus* base. O ALL foi constituído com o conteúdo de todas as memórias de tradução com o par de línguas português-inglês do Departamento de Línguas da ALERT. Ainda neste âmbito, poderia ter sido útil utilizar um *corpus* de domínio geral que resultasse de traduções escritas, ao

contrário do Europarl, que é composto por transcrições traduzidas das sessões do Parlamento Europeu.

Igualmente útil poderia ter sido a utilização do *script* disponibilizado pelo Dr. Hilário Fontes para a atualização da ortografia dos resultados do MMM para o Acordo Ortográfico de 1990. A escassez de tempo aliada à curva de aprendizagem necessária para utilizar os outros *scripts* impediram que este recurso fosse utilizado. Também por esse facto, e pelo facto de os *corpora* com português seguirem maioritariamente a norma do Acordo Ortográfico de 1945, não se considerou falha ou erro sempre que uma palavra seguia a ortografia de 1945 e não a de 1990. O facto de as traduções do Google apresentarem tanto a variante ortográfica do português do Brasil como do português de Portugal e não unicamente a do português de Portugal, como aconteceu com o MMM, não teve qualquer influência na avaliação da qualidade, uma vez que esta diferença se restringiu à ortografia e, como tal, não influenciou o juízo de gramaticalidade dos segmentos traduzidos, nem deturpou a informação transferida do TP para o TC. Além disso, o Google não permite selecionar variantes linguísticas no seu sistema de TAE. Portanto, de forma a ser possível comparar resultados de ambos os sistemas de TAE, optou-se por ignorar questões ortográficas que tivessem tanto a ver com acordos ortográficos como com variantes do português. Num estudo futuro, convém garantir que ambos os sistemas estão preparados para devolver resultados que respeitem as mesmas regras ortográficas e a mesma variante linguística. Como tal, será importante garantir que os sistemas usados são treinados com *corpora* selecionados para o efeito e não são sistemas fechados que não permitam o controlo do conteúdo com que os motores de tradução são treinados (como acontece com o Google Tradutor).

Além da importância já referida que conhecimentos básicos em termos de estatística poderiam ter tido para compreender melhor muitos dos conceitos teóricos que sustentam os componentes informáticos de um sistema de TAE, esses conhecimentos teriam sido também importantes para uma interpretação mais aprofundada dos resultados quantitativos e qualitativos resultantes da avaliação das traduções. Apesar de termos tentado processar o melhor possível toda a informação disponibilizada pela avaliação manual, penso que se tivéssemos mais conhecimentos acerca de estatística adaptada à investigação, poderíamos ter conseguido retirar mais conclusões e mais aprofundadas. Como tal, penso que seria útil incluir um módulo de análise estatística e interpretação de dados, no âmbito dos seminários de Metodologia da Investigação.

Este projeto, além de nos ter possibilitado adquirir conhecimentos teóricos e práticos acerca de sistemas de TAE, nomeadamente do MMM, permitiu-nos perceber que é cada vez mais necessário que todos aqueles que trabalham com a língua adquiram competências tecnológicas que lhes possibilitem utilizar as ferramentas que se encontram à sua disposição para realizar o seu trabalho de forma mais eficaz e eficiente. Contudo, ferramentas como sistemas de TA são concebidas e construídas essencialmente por pessoas com formação informática. Caso os tradutores ou linguistas pretendam intervir de forma ativa no desenvolvimento de ferramentas informáticas que sirvam as suas atividades é necessário que adquiram conhecimentos na área das ciências informáticas, caso contrário, essas ferramentas dificilmente cumprirão na plenitude o propósito para que foram criadas.

Bibliografia

- ALPAC. 1966. *Languages and machines: computers in translation and linguistics*. Washington, D.C.: National Academy of Sciences, National Research Council, 124. Acedido em setembro de 2014. http://www.nap.edu/openbook.php?record_id=9547.
- Axelrod, A., H. Xiaodong, e G. Jianfeng. 2011. "Domain adaptation via pseudo in-domain data selection." *EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA (EUA): Association for Computational Linguistics. 355-362.
- Bar-Hillel, Y. 1951. "The state of machine translation in 1951." *American Documentation* 2 229-237. *apud* Hutchins, John. 1986. *Machine Translation: past, present, future*. New York: Halsted Press, p. 10
- Bose, R. 2004. "Natural Language Processing: Current State and Future Directions." *International Journal of the Computer, the Internet and Management*, 1-11. Acedido em 6 de setembro de 2014. http://ijcim.th.org/past_editions/2004V12N1/ijcimvol12n1_article1.pdf.
- Brown, F. Peter, Vicent J. Della Pietra, Stephen A. Della Pietra, e Robert L. Mercer. 1993. "The mathematics of statistical machine translation: parameter estimation." *Journal of Computational Linguistics – Special issue on using large corpora*, junho: 263-311. Acedido em setembro de 2014. <http://dl.acm.org/citation.cfm?id=972474>.
- Cardoso, Nuno, e Alberto Simões. 2014. *Tudo o que você queria saber sobre configuração de codificações em português*. julho. <http://www.linguateca.pt/codificacoes/>.
- Chatzitheodorou, Konstantinos, e Stamatis Chatzistamatis. 2010. "COSTA MT Evaluation Tool: An Open Toolkit for Human Machine Translation Evaluation." *The Prague Bulletin of Mathematical Linguistics* (100): 83-89. Acedido em setembro de 2014. <https://ufal.mff.cuni.cz/pbml/100/art-chatzitheodorou-chatzistamatis.pdf>.
- Chérargui, M. 2012. "Theoretical Overview of Machine Translation." *International conference on Web and Information Technologies (ICWIT)*. 160-169. Acedido em setembro de 2014. <http://ceur-ws.org/Vol-867/Paper17.pdf>.
- Collins, Michael. s.d. *Machine Translation*. Powerpoint. Columbia University. Acedido em julho de 2014. <http://www.cs.columbia.edu/~cs4705/lectures/mt1.pdf>.
- Directorate-General for Translation. 2014. *English Style Guide: A handbook for authors and translators in the European Commission*. Comissão Europeia. Acedido em julho de 2014.

http://ec.europa.eu/translation/english/guidelines/documents/styleguide_english_dgt_en.pdf.

Doherty, S., e Federico G. 2013. *Understanding and Implementing Effective Translation Quality Evaluation Techniques*. Powerpoint. Acedido em agosto de 2014.
<http://www.qt21.eu/launchpad/sites/default/files/QTLP%20GALA%20Webinar%203.pdf>.

España, Cristina. 2011. *Un primer cop d'ull a la Traducció Automàtica Estadística*. Powerpoint.

Euromatrix. 2007. *Survey of Machine Translation Evaluation*. Comissão Europeia. Acedido em agosto de 2014.
http://www.euromatrix.net/deliverables/Euromatrix_D1.3_Revised.pdf.

Federman, Christian. 2012. "Appraise: an Open-Source Toolkit for Manual Evaluation of MT output." *35th German Conference on Artificial Intelligence*. Saarbrücken, Alemanha. Acedido em setembro de 2014.
<http://www.dfki.de/KI2012/PosterDemoTrack/ki2012pd27.pdf>.

Hutchins, John. 2003a. "Commercial Systems: state of the art." Em *Computers and translation: a translator's guide*, editado por Harold Somers, 161-174. Amsterdam: John Benjamins. Acedido em setembro de 2014.
<http://www.hutchinsweb.me.uk/Somers-2003.pdf>.

—. 2011. "Machine Translation systems and applications: past and present." *John Hutchins – Publications on machine translation, computer-based translation technologies, linguistics and other topics*. Acedido em 05 de julho de 2012.
<http://www.hutchinsweb.me.uk/ImpColl-2011-ppt.pdf>.

—. 2010. "Machine Translation: a concise history." *Journal of Translation Studies*, vol.13, nos.1-2 (2010). *Special issue: The teaching of computer-aided translation*. Chinese University of Hong Kong. 29-70. Acedido em 05 de julho de 2012. <http://www.hutchinsweb.me.uk/CUHK-2006.pdf>.

Hutchins, John. 2003b. "Machine Translation: General Overview." Em *The Oxford Handbook of Computational Linguistics*, editado por Ruslan Mitkov, 501-511. Oxford: Oxford University Press. Acedido em setembro de 2014.
<http://www.hutchinsweb.me.uk/Mitkov-2003.pdf>.

Hutchins, John. 2006. *Machine Translation: History*. Vol. 7, em *Encyclopedia of Language & Linguistics*, editado por Keith Brown, 375-383. Oxford: Elsevier.

—. 1986. *Machine Translation: past, present, future*. New York: Halsted Press.

—. 2009. "Multiple uses of machine translation and computerized translation tools." *ISMTCL: International Symposium on Data Mining and Sense Mining, Machine Translation and Controlled Languages, and their application to emergencies*

and safety critical domains. Besançon, France: Presses universitaires de Franche-Comté. 13-20.

- Jurafsky, Daniel, e James H. Martin. 2000. *Speech and Language Processing*. Pearson Education Inc. *apud* Ramanathan, A. s.d. “Statistical Machine Translation: PhD seminar report.” Department of Computer Science and Engineering, Indian Institute of Technology, Bombaim. p. 13. Acedido em julho de 2014. http://202.141.152.9/clir/papers/statistical_mt.pdf.
- Knight, Kevin, e Phillip Koehn. s.d. *What’s new in statistical machine translation*. Powerpoint. Acedido em agosto de 2014. <http://homepages.inf.ed.ac.uk/pkoehn/publications/tutorial2003.pdf>.
- Koehn, Philipp. 2005. “Europarl: A Parallel Corpus for Statistical Machine Translation.” *MT Summit*. 79-86. Acedido em julho de 2014. <http://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf>.
- Koehn, Phillip. 2010. *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Léon, J. 1997. “Les premières machines à traduire (1948–1960) et la filiation cybernétique.” *Bulag* 22, 9-33. *apud* Hutchins, John. 2006. *Machine Translation: History*. Vol. 7, em *Encyclopedia of Language & Linguistics*, editado por Keith Brown, 375-383. Oxford: Elsevier. p. 377.
- Lopez, A. 2008. “Statistical Machine Translation.” *ACM Computing Surveys* 3: 49.
- Machado, M. José, e Hilário Fontes. 2011. “Avaliação do sistema de tradução automática Moses no Departamento de Língua Portuguesa da DGT utilizando os scripts Moses for Mere Mortals — estudo de caso inglês-português.” *a folha Boletim da língua portuguesa nas instituições europeias*, 12-23. Acedido em junho de 2014. http://ec.europa.eu/translation/portuguese/magazine/documents/folha37_pt.pdf.
- Machado, M. José, Hilário Fontes, e João Rosas. 2010. “Sistema de tradução automática Moses e... Moses for Mere Mortals.” *a folha Boletim da língua portuguesa nas instituições europeias*. Acedido em junho de 2014. http://ec.europa.eu/translation/portuguese/magazine/documents/folha32_pt.pdf.
- Machado, M. José, Hilário Fontes, e João Rosas. s.d. *Moses for Mere Mortals – Tutorial*. Acedido em junho de 2014. <http://moses-for-mere-mortals.googlecode.com/files/Help-Short-Tutorial.doc>.
- Nagao, M. 1984. “A framework of a mechanical translation between Japanese and English by analogy principle.” Em *Artificial And Human Intelligence*, editado por A. Elithorn e R. Banerji. Elsevier Science Publishers. B.V. Acedido em setembro de 2014. <http://mt-archive.info/Nagao-1984.pdf>.

- Ostler, Nicholas. 2010. *The Last Lingua Franca: The Rise and Fall of World Languages*. London: Penguin Books.
- Papineni, K., S. Roukos, T Ward, e W. Zhu. 2002. “BLEU: a Method for Automatic Evaluation of Machine Translation.” *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Filadélfia. 311-318.
- Ramanathan, A. s.d. “Statistical Machine Translation: PhD seminar report.” Department of Computer Science and Engineering, Indian Institute of Technology, Bombaim. Acedido em julho de 2014.
http://202.141.152.9/clir/papers/statistical_mt.pdf.
- Shannon, C. E. 1948. “A mathematical theory of communication.” *The Bell System Technical Journal* 27: 379–423, 623–656. Acedido em julho de 2014.
<http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>.
- Simões, A. 2009. *Abordagens na Tradução Automática*. Powerpoint. Acedido em setembro de 2014.
http://web.letras.up.pt/bhsmaia/EDV/apresentacoes/Simoes_Arquitectures_MT.pdf.
- Somers, Harold. 2003. “Machine Translation: latest developments.” Em *The Oxford Handbook of Computational Linguistics*, editado por Ruslan Mitkov, 512-528. Oxford: Oxford University Press. Acedido em setembro de 2014.
<http://personalpages.manchester.ac.uk/staff/harold.somers/Mitkov-book-chapter.pdf>.
- Specia, L. 2013. “Statistical Machine Translation.” Em *Emerging Applications of Natural Language Processing: Concepts and New Research.*, de S., Naskar, S.K., Ekbal, A. Bandyopadhyay, 74-109. IGI Global.
- Specia, L. 2010. *Tradução Automática Estatística: Tutorial PROPOR*. Powerpoint. Acedido em 23 de setembro de 2014. http://pers-www.wlv.ac.uk/~in1316/resources/statisticalmachinetranslation_propor.pdf.
- Stein, D. 2013. “Machine Translation: Past, Present and Future.” *Translation: Computation, Corpora, Cognition (TC3): Special Issue on Language Technologies for a Multilingual Europe*. Acedido em agosto de 2014.
<http://www.t-c3.org/index.php/t-c3/article/view/28>.
- Tiedemann, Jörg. 2009. *News from OPUS – A Collection of Multilingual Parallel Corpora with Tools and Interfaces*. Vol. V, em *Recent Advances in Natural Language Processing*, editado por N. Nicolov, K. Bontcheva, G. Angelova e R. Mitkov, 237-248. Amesterdão/Filadélfia: John Benjamins.
- . 2012. “Parallel Data, Tools and Interfaces in OPUS.” *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*.

- Tripathi, S., e J. K. Sarkhel. 2010. “Approaches to machine translation.” *Annals of Library and Information Studies*, dezembro: 388-393. Acedido em setembro de 2014. <http://nopr.niscair.res.in/handle/123456789/11057>.
- Weaver, W. 1949. “Memo on machine translation.” *MT-archives*. Editado por The Rockefeller Foundation. Acedido em 16 de agosto de 2014. <http://www.mt-archive.info/Weaver-1949.pdf>.
- Zhang, J, X. Wu, I. Calixto, A. H. Vahid, X Zhang, A. Way, e Q. Liu. 2014. “Experiments in Medical Translation Shared Task at WMT 2014.” *Proceedings of the Ninth Workshop on Statistical Machine Translation*. Baltimore, Maryland (EUA): Association for Computational Linguistics. 260–265. Acedido em 19 de setembro de 2014. <http://www.statmt.org/wmt14/pdf/W14-3332.pdf>.

Referências web

- ALERT Life Sciences Computing. s.d. Acedido em setembro de 2014. <http://www.alert-online.com/>.
- American Translators Association. s.d. *American Translators Association*. Acedido em setembro de 2014. http://www.atanet.org/certification/aboutexams_error.php.
- Chatzitheodorou, Konstantinos, e Stamatis Chatzistamatis. 2010. “COSTA MT Evaluation Tool: An Open Toolkit for Human Machine Translation Evaluation.” *The Prague Bulletin of Mathematical Linguistics* (100): 83-89. Acedido em setembro de 2014. <https://ufal.mff.cuni.cz/pbml/100/art-chatzitheodorou-chatzistamatis.pdf>.
- s.d. *Frequently asked questions about Computational Linguistics*. Acedido em setembro de 2014. http://web.archive.org/web/20100805080705/http://www.ling.su.se/DaLi/cl_faq/index.htm.
- Google Inc. s.d. *About Google Translate*. Acedido em setembro de 2014. http://translate.google.pt/about/intl/pt-PT_ALL/.
- . s.d. *Google Play*. Acedido em setembro de 2014. <https://play.google.com/store/apps/>.
- . s.d. *Google Tradutor*. Acedido em setembro de 2014. <https://translate.google.pt/>.
- . s.d. *NLP at Google Research*. Acedido em setembro de 2014. <http://research.google.com/pubs/NaturalLanguageProcessing.html>.
- . s.d. *Translator Toolkit*. Acedido em junho de 2014. https://translate.google.com/toolkit/list?hl=pt_PT#translations/active.

s.d. *Lingueca*. Acedido em setembro de 2014. <http://www.lingueca.pt/>.

s.d. *Moses for Mere Mortals*. Acedido em setembro de 2014.
<https://code.google.com/p/moses-for-mere-mortals/> .

s.d. *Moses Statistical Machine Translation System*. Acedido em setembro de 2014.
<http://www.statmt.org/moses/>.

National Institute of Standards and Technology. s.d. *About NIST*. Acedido em setembro de 2014. http://www.nist.gov/public_affairs/nandyou.cfm.

OKpedia. s.d. *Mainframe*. Acedido em junho de 2014.
<http://pt.okpedia.com/mainframe>.

s.d. *Opus... The open parallel corpus*. Acedido em setembro de 2014.
<http://opus.lingfil.uu.se/>.

Porto Editora. s.d. *Infopédia*. Acedido em junho de 2014. <http://www.infopedia.pt>.

Princeton University. s.d. *Stochastic process*. Acedido em junho de 2014.
http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Stochastic_process.html.

RAND. s.d. *About the RAND Corporation*. Acedido em setembro de 2014.
<http://www.rand.org/about.html>.

Trinity University. s.d. *Bayes' Theorem*. Acedido em setembro de 2014.
<http://www.trinity.edu/cbrown/bayesWeb/>.

Ubuntu. s.d. *Ubuntu*. Acedido em maio de 2014. <http://old-releases.ubuntu.com/releases/10.04.3/>.

Wikipedia. s.d. *Wikipedia*. Acedido em junho de 2014. <http://en.wikipedia.org>.

Anexos

i. “Instruções de Alta” – Original e tradução de referência

	Original_ENUS	Tradução_PT
1	Chronic hypertension	Hipertensão arterial crônica
2	Chronic hypertension is a long-term condition in which your blood pressure (BP) is higher than normal.	A hipertensão arterial crônica é uma doença de longa duração em que a pressão arterial (PA) é superior ao normal.
3	Your BP is the force of your blood moving against the walls of your arteries.	A pressão arterial é a força da movimentação do sangue contra as paredes das artérias.
4	Normal blood pressure is less than 120/80.	Os valores normais da pressão arterial situam-se abaixo de 120/80 mmHg.
5	Prehypertension is BP between 120/80 and 139/89.	A pré-hipertensão arterial situa-se entre 120/80 mmHg e 139/89 mmHg.
6	High blood pressure is 140/90 or higher.	A hipertensão arterial situa-se a partir de 140/90 mmHg.
7	Medicines that may be ordered for you	Medicação que lhe pode ser receitada
8	- Blood pressure medicine is given to lower your blood pressure.	- A medicação para a hipertensão é administrada para baixar a pressão arterial.
9	A controlled blood pressure helps protect your organs, such as your heart, lungs, brain, and kidneys.	Uma pressão arterial controlada ajuda a proteger os órgãos, como o coração, pulmões, cérebro e rins.
10	You may need more than 1 type of blood pressure medicine.	Poderá precisar de mais do que um tipo de medicamento para a pressão arterial.
11	Take your blood pressure medicine exactly as directed.	Tome a medicação para a pressão arterial exatamente como prescrito.
12	- Diuretics help decrease extra fluid that collects in your body.	- Os diuréticos ajudam a diminuir os líquidos em excesso acumulados no seu organismo.
13	This will help lower your BP.	Isto irá ajudar a baixar a pressão arterial.
14	You may urinate more often while you take this medicine.	É possível que urine com maior frequência enquanto estiver a tomar este medicamento.
15	- Take your medicine as directed.	- Tome a medicação conforme indicado.
16	Call your healthcare provider if you think your medicine is not helping or if you have side effects.	Contacte o seu profissional de saúde se pensa que o medicamento não está a ajudar ou se tiver efeitos secundários.
17	Tell him if you are allergic to any medicine.	Informe se tiver alergia a algum medicamento.
18	Keep a list of the medicines, vitamins, and herbs you take.	Mantenha consigo uma lista dos medicamentos, vitaminas ou produtos de ervanária que esteja a tomar.

19	Include the amounts, and when and why you take them.	Inclua as quantidades, quando e por que razão os toma.
20	Bring the list or the pill bottles to follow-up visits.	Leve a lista ou as embalagens dos seus medicamentos para as consultas de seguimento.
21	Carry your medicine list with you in case of an emergency.	Leve a lista de medicação consigo para casos de emergência.
22	Pulmonary hypertension	Hipertensão pulmonar
23	Pulmonary hypertension, or PH, is increased pressure in the pulmonary arteries.	A hipertensão pulmonar (HP) consiste no aumento da pressão nas artérias pulmonares.
24	These arteries carry blood from your heart to your lungs to pick up oxygen.	Estas artérias transportam o sangue do coração aos pulmões para recolher oxigénio.
25	PH causes symptoms such as shortness of breath during routine activity (for example, climbing two flights of stairs), tiredness, chest pain, and a racing heartbeat.	A HP provoca sintomas, tais como falta de ar durante atividades de rotina (por exemplo, subir dois lanços de escadas), fadiga, dor torácica e ritmo cardíaco acelerado.
26	As the condition worsens, its symptoms may limit all physical activity.	À medida que a doença vai progredindo, os seus sintomas podem limitar qualquer tipo de atividade física.
27	To understand PH, it helps to understand how your heart and lungs work.	Para entender a HP, ajuda compreender primeiro o funcionamento do coração e dos pulmões.
28	Your heart has two sides, separated by an inner wall called the septum.	O coração tem dois lados, separados por uma parede interior designada septo.
29	Each side of your heart has an upper and lower chamber.	Cada lado do coração possui uma câmara superior e uma inferior.
30	The lower right chamber of your heart, the right ventricle, pumps blood to your pulmonary arteries.	A câmara inferior direita do coração, o ventrículo direito, bombeia sangue para as artérias pulmonares.
31	The blood then travels to your lungs, where it picks up oxygen.	O sangue é encaminhado para os pulmões, onde recolhe oxigénio.
32	The upper left chamber of your heart, the left atrium, receives the oxygen-rich blood from your lungs.	A câmara superior esquerda do coração, a aurícula esquerda, recebe sangue rico em oxigénio dos pulmões.
33	The blood is then pumped into the lower left chamber of your heart, the left ventricle.	O sangue é bombeado para dentro da câmara inferior esquerda do coração, o ventrículo esquerdo.
34	From the left ventricle, the blood is pumped to the rest of your body through an artery called the aorta.	A partir do ventrículo esquerdo, o sangue é bombeado para o resto do corpo através de uma artéria denominada aorta.
35	PH begins with inflammation and changes in the cells that line your pulmonary arteries.	A HP começa com a inflamação e alterações nas células que revestem as artérias pulmonares.
36	Other factors also can affect the pulmonary arteries and cause PH.	Outros fatores poderão igualmente afetar as artérias pulmonares e causar HP.

37	For example, the condition may develop if:	Por exemplo, a doença pode desenvolver-se se:
38	- The walls of the arteries tighten.	- As paredes das artérias estreitam.
39	- The walls of the arteries are stiff at birth or become stiff from an overgrowth of cells.	- As paredes das artérias estiverem rígidas à nascença ou ficarem rígidas com a proliferação de células.
40	- Blood clots form in the arteries.	- Formarem coágulos sanguíneos nas artérias.
41	- These changes make it hard for your heart to push blood through your pulmonary arteries and into your lungs.	- Estas alterações dificultam a tarefa do coração de bombear sangue através das artérias pulmonares para os pulmões.
42	As a result, the pressure in your arteries rises.	Como resultado, a pressão nas artérias aumenta.
43	Also, because your heart is working harder than normal, your right ventricle becomes strained and weak.	Além disso, uma vez que o coração se está a esforçar mais do que o normal, o ventrículo direito fica cansado e fraco.
44	Your heart may become so weak that it can't pump enough blood to your lungs.	O coração pode ficar tão debilitado que não consiga bombear sangue suficiente para os pulmões.
45	This causes heart failure.	Isto provoca insuficiência cardíaca.
46	Heart failure is the most common cause of death in people who have PH.	A insuficiência cardíaca é a causa de morte mais frequente em doentes com HP.
47	PH is divided into five groups based on its causes and treatment options.	A HP está dividida em cinco grupos com base nas suas causas e opções terapêuticas.
48	In all groups, the average pressure in the pulmonary arteries is 25 mmHg or higher.	A pressão média nas artérias pulmonares, em todos os grupos, é de 25 mmHg ou superior.
49	The pressure in normal pulmonary arteries is 8–20 mmHg at rest.	A pressão normal nas artérias pulmonares é de 8–20 mmHg em repouso.
50	(The mmHg is millimeters of mercury—the units used to measure blood pressure.)	(mmHg corresponde a milímetros de mercúrio — a unidade utilizada para medir a pressão arterial.)
51	PH has no cure.	A HP não tem cura.
52	However, research for new treatments is ongoing.	Contudo, estão a ser investigados novos tratamentos.
53	The earlier PH is treated, the easier it is to control.	Quanto mais cedo a HP for tratada, mais fácil se tornará o seu controlo.
54	Treatments include medicines, procedures, and other therapies.	Os tratamentos incluem medicação, procedimentos, entre outras terapias.
55	These treatments can relieve PH symptoms and slow the progress of the disease.	Estes tratamentos podem aliviar os sintomas da HP e atrasar a evolução da doença.
56	Lifestyle changes also can help control symptoms.	Alterações ao estilo de vida também poderão ajudar a controlar os sintomas.
57	Lifestyle changes	Alterações ao estilo de vida

58	- Monitor your weight: weigh yourself at the same time each day and record your weight in a diary or on a calendar.	- Vigiar o seu peso: pese-se todos os dias à mesma hora e registe o seu peso num diário ou calendário.
59	- Lose weight if you are overweight (follow your doctor's recommendations).	- Perca peso, se tiver excesso de peso (siga as instruções do seu médico).
60	- Have annual health check-ups with your primary care physician.	- Faça um exame geral de rotina todos os anos com o seu médico de família.
61	- Activity guidelines: restrict lifting, pushing, or shoving to less than 20 pounds, since these activities increase the pressure in your arteries and lungs; walking is important to keep your muscles strong and improve circulation.	- Indicações para atividades: não levante ou empurre pesos superiores a 9 kg, uma vez que estas atividades aumentam a pressão nas artérias e pulmões. Caminhar é importante para manter os músculos fortes e melhorar a circulação.
62	However, more strenuous forms of aerobic exercise should not be done regularly since your lungs may not be able to keep up with the oxygen demands placed on your body during these activities.	Não deve realizar exercícios aeróbicos intensos regularmente, uma vez que os pulmões poderão não acompanhar as necessidades de oxigénio do corpo durante estas atividades.
63	Any activities that cause shortness of breath, dizziness, or chest pain should be stopped immediately.	Quaisquer atividades que provoquem falta de ar, vertigem ou dor no peito devem ser imediatamente suspensas.
64	Contact your doctor if any of these symptoms occur.	Contacte o médico caso surja algum destes sintomas.
65	- Do not smoke or chew tobacco.	- Não fume ou mastigue tabaco.
66	- Avoid or reduce alcohol intake.	- Evite ou reduza o consumo de álcool.
67	- Avoid pregnancy; use contraceptives to prevent pregnancy.	- Evite a gravidez; use contraceptivos para evitar a gravidez.
68	- Maintain close contact with your health care providers.	- Mantenha contacto próximo com os seus profissionais de saúde.
69	When to call your doctor?	Quando contactar o médico?
70	- Weight gain — 2 pounds in one day or 5 pounds in one week	- Aumento de peso - 1 kg num dia ou 2,5 kg numa semana.
71	- Swelling in your ankles, legs, or abdomen that has become worse	- Inchaço dos tornozelos, pernas ou abdómen que tenha piorado.
72	- Shortness of breath that has become worse, especially if you wake up short of breath	- Falta de ar que tenha piorado, especialmente se já tiver acordado com falta de ar
73	- Extreme fatigue or decreased tolerance for activity	- Fadiga extrema ou menor tolerância a atividade
74	- A respiratory infection or a cough that has become worse	- Infecção respiratória ou tosse que tenha piorado
75	- Fast heart rate (around 120 beats per minute)	- Frequência cardíaca elevada (cerca de 120 batimentos por minuto)

76	- Episodes of chest pain or discomfort with exertion that are not relieved with rest	- Episódios de dor ou desconforto no peito provocados por esforço e que não diminuem com repouso
77	- Difficulty breathing with regular activities or at rest	- Dificuldade em respirar com atividades normais ou em repouso
78	- Restlessness, confusion	- Inquietação, confusão
79	- Constant dizziness or lightheadedness	- Tonturas ou sensação de desfalecimento constante
80	- Nausea or poor appetite	- Náuseas ou falta de apetite
81	When to go to the emergency department?	Quando deverá dirigir-se ao Serviço de Urgência?
82	- Hickman catheter complications with intravenous Flolan, such as infection, catheter displacement/leakage of solution or blood, IV pump malfunction	- Complicações com cateter de Hickman para perfusão de Flolan, tais como infecções, deslocação de cateter/perda de soro ou sangue, avaria da bomba infusora
83	- New chest pain or discomfort that is severe, unexpected, and is accompanied by shortness of breath, sweating, nausea, or weakness	- Surgimento de dor ou desconforto torácico intenso, inesperado e acompanhado de falta de ar, suores, náuseas ou fraqueza
84	- Fast, sustained heart rate (more than 120 to 150 beats per minute) — especially if you are short of breath	- Frequência cardíaca acelerada e constante (mais de 120 a 150 batimentos por minuto) — especialmente se acompanhada de falta de ar
85	- Shortness of breath that is NOT relieved by rest	- Falta de ar que não seja aliviada pelo repouso
86	- Sudden weakness or paralysis in your arms or legs	- Fraqueza ou paralisia súbita nos braços ou pernas
87	- Sudden onset of severe headache	- Início súbito de dores de cabeça intensas
88	- Fainting spells with loss of consciousness	- Episódios de desmaio com perda de consciência
89	Preeclampsia and Eclampsia	Pré-eclâmpsia e eclâmpsia
90	Preeclampsia (toxemia) and eclampsia are conditions that can develop during week 20 or later of your pregnancy.	Pré-eclâmpsia (toxemia) e eclâmpsia são condições que podem surgir às 20 ou mais semanas de gestação.
91	With preeclampsia, you have high blood pressure and protein in your urine.	Com pré-eclâmpsia os seus valores da pressão arterial e de proteínas na urina são elevados.
92	Preeclampsia progresses to eclampsia if you have a seizure.	A pré-eclâmpsia evolui para eclâmpsia se tiver uma convulsão.
93	This does not include women with a known seizure disorder.	Isto não inclui mulheres com problemas de epilepsia.
94	These conditions can create mild to life-threatening health problems for you and your unborn baby.	Estas condições podem dar origem a problemas de saúde ligeiros a potencialmente fatais para si e para o seu filho.
95	You will need tests 1 to 2 times a week to check your condition.	Necessitará de realizar exames 1 a 2 vezes por semana, para controlar a sua doença.

96	Tests include blood pressure checks, urine and blood tests, and fetal monitoring.	Estes exames incluem a medição da tensão arterial, análises à urina e ao sangue e monitorização fetal.
97	Rest:	Repouso:
98	- Your obstetrician may tell you to rest more often if you have mild symptoms of preeclampsia.	- O seu médico obstetra pode dizer-lhe para descansar com maior frequência se tiver sintomas ligeiros de pré-eclâmpsia.
99	You may need total bedrest if you have more severe symptoms.	Pode necessitar de repouso total na cama se tiver sintomas mais graves.
100	Try to lie on your left side.	Procure deitar-se do seu lado esquerdo.

ii. “Instruções de Alta” – Avaliação da tradução do Moses for Mere Mortals (ALL)

	ORIGINAL_EN	MMM_PT (ALL)	Gr a	Ter	Po	Om	Tra	Mai	Coe
1	Chronic hypertension	Crónica Hipertensão	1						
2	Chronic hypertension is a long-term condition in which your blood pressure (BP) is higher than normal.	Crónica Hipertensão é uma condição de longa duração nos quais a Pressão arterial (PA) é maior do que o normal.	1					1	
3	Your BP is the force of your blood moving against the walls of your arteries.	A BP é a força do sangue o movimento contra a parede do artérias.	3						1
4	Normal blood pressure is less than 120/80.	Sem tensão arterial é menos de 120/80.					3		1
5	Prehypertension is BP between 120/80 and 139/89.	Prehypertension é entre PA 120/80 e 139/89.	2				2		
6	High blood pressure is 140/90 or higher.	Hipertensão está 140/90 ou superior.	2						
7	Medicines that may be ordered for you	Medicamentos que podem ser requisitadas por si					3		
8	- Blood pressure medicine is given to lower your blood pressure.	- Pressão arterial Medicina é dada a baixar a Pressão arterial.	1				3	1	
9	A controlled blood pressure helps protect your organs, such as your heart, lungs, brain, and kidneys.	Um Encerramento controlado Pressão arterial. Ajuda a proteger os órgãos, tais como o coração, pulmões, cérebro e rins.			2		3		
10	You may need more than 1 type of blood pressure medicine.	Poderá ser necessário recorrer a mais do que um tipo de tensão arterial.				3			1
11	Take your blood pressure medicine exactly as directed.	Tome a sua Pressão arterial a medicação exatamente conforme as instruções.	2					1	
12	- Diuretics help decrease extra fluid that collects in your body.	- diuréticos ajudar a Diminuir extra líquido que faz no seu corpo.	1				3	1	
13	This will help lower your BP.	Isto irá ajudar a baixar a Pressão arterial.						1	
14	You may urinate more often while you take this medicine.	Poderá urinar com mais frequência durante a tomar Esta medicação.	2					1	
15	- Take your medicine as directed.	- Tome a medicação conforme as instruções.							
16	Call your healthcare provider if you think your medicine is not helping or if you have side effects.	Contacte o seu médico se a medicação está aumentando ou se sentir efeitos secundários.				3	3		
17	Tell him if you are allergic to any medicine.	Irá confirmar se tem alergia a algum medicamento.					3		
18	Keep a list of the medicines, vitamins, and herbs you take.	Mantenha uma lista com as vitaminas, produtos de ervanária ou dar.				2	3		
19	Include the amounts, and when and why you take them.	Incluem os montantes e, quando e das razões que os levam.	2		1		3		

20	Bring the list or the pill bottles to follow-up visits.	Acompanhado da lista e as embalagens dos seus medicamentos para as consultas de acompanhamento.	1			1			
21	Carry your medicine list with you in case of an emergency.	Consigo a sua medicação consigo lista em caso de emergência.	3			1			
22	Pulmonary hypertension	Hipertensão pulmonar							
23	Pulmonary hypertension, or PH, is increased pressure in the pulmonary arteries.	Hipertensão pulmonar ou pH, será aumentada Pressão pulmonares artérias.	2	2			3	1	
24	These arteries carry blood from your heart to your lungs to pick up oxygen.	Essas artérias Pegue sangue do seu coração para os pulmões pegar oxigénio.					3	1	
25	PH causes symptoms such as shortness of breath during routine activity (for example, climbing two flights of stairs), tiredness, chest pain, and a racing heartbeat.	PH causam sintomas como Dispneia durante as atividades (por exemplo, dois subir escadas de voos), cansaço, dores no peito, e aceleração frequência cardíaca.	2			2	2	1	
26	As the condition worsens, its symptoms may limit all physical activity.	Pois a condição aumentam, os sintomas possam reduzir a toda a actividade Física.	3			1	3	1	
27	To understand PH, it helps to understand how your heart and lungs work.	Perceber pH, ajuda a perceber como o seu coração e pulmões, pertencendo		2	1		3		
28	Your heart has two sides, separated by an inner wall called the septum.	Coração com dois lados, separados por uma parede denominadas septo.	1			1	2		
29	Each side of your heart has an upper and lower chamber.	Cada lado do coração tem superiores e inferiores câmara.	2						
30	The lower right chamber of your heart, the right ventricle, pumps blood to your pulmonary arteries.	Canto inferior direito da câmara do coração, do ventrículo direito, aspiradores de sangue para a tua pulmonares artérias.	2				3		
31	The blood then travels to your lungs, where it picks up oxygen.	Sangue, viagens aos seus pulmões, onde picks Up oxigénio.				1	3		
32	The upper left chamber of your heart, the left atrium, receives the oxygen-rich blood from your lungs.	Canto superior esquerdo da câmara do coração, do Átrio esquerdo, é dada a oxygen-rich sangue dos seus pulmões.	3				3	1	1
33	The blood is then pumped into the lower left chamber of your heart, the left ventricle.	O sangue é então pumped no canto inferior esquerdo câmara do coração, do ventrículo esquerdo.	3				3		
34	From the left ventricle, the blood is pumped to the rest of your body through an artery called the aorta.	Do ventrículo esquerdo, o sangue é pumped para o resto da sua corporais através de um chamado Artéria aorta.	2				3	1	
35	PH begins with inflammation and changes in the cells that line your pulmonary arteries.	PH começa a inflamação e alterações nas células que a linha pulmonares artérias.	1			1	3		
36	Other factors also can affect the pulmonary arteries and cause PH.	Outros factores também podem afectar o pulmonares artérias e provocam pH.	2	2					
37	For example, the condition may develop if:	Por exemplo, a condição poderá gerar-se se:	1						

38	- The walls of the arteries tighten.	- o sabão nas paredes das artérias apertar .	1				3	1	
39	- The walls of the arteries are stiff at birth or become stiff from an overgrowth of cells.	- o sabão nas paredes das artérias rígidos à nascença ou tornar-se Rígida em qualquer overgrowth celulares .	1			2	3	1	
40	- Blood clots form in the arteries.	- formar coágulos sanguíneos.	1					1	
41	- These changes make it hard for your heart to push blood through your pulmonary arteries and into your lungs.	- estas alterações ser difícil para o seu coração para se sanguíneo através da sua pulmonar artérias e para os pulmões.	1			3	3	1	
42	As a result, the pressure in your arteries rises.	Como resultado, a Pressão nas artérias rises .					3	1	
43	Also, because your heart is working harder than normal, your right ventricle becomes strained and weak.	Também, pois o seu coração está a funcionar mais tempo que o normal, ventrículo direito fica strained e fracos .	1			1	3		
44	Your heart may become so weak that it can't pump enough blood to your lungs.	Seu coração para que o processo possa ficar fraco não conseguem bomba de sangue suficiente para os pulmões.	3			3			
45	This causes heart failure.	Isto Insuficiência cardíaca.				3		1	
46	Heart failure is the most common cause of death in people who have PH.	Insuficiência cardíaca é a causa mais comum para mortos nas pessoas que têm pH .	1	2			2		
47	PH is divided into five groups based on its causes and treatment options.	PH está dividido em cinco grupos com base nas suas causas e opções de tratamento.	1				2		
48	In all groups, the average pressure in the pulmonary arteries is 25 mmHg or higher.	Nos grupos, a média Pressão pulmonares artérias muito comum: 25 mmHg ou superior.	1			1	2	1	
49	The pressure in normal pulmonary arteries is 8–20 mmHg at rest.	A Pressão pulmonar Sem alterações artérias é 8 – 20 mmHg em repouso.	3				1	1	
50	(The mmHg is millimeters of mercury—the units used to measure blood pressure.)	(mmHg são millimeters de Mercúrio — a unidade utilizada para medir a Pressão arterial.)					1	1	
51	PH has no cure.	PH não possui cure .	1				2		
52	However, research for new treatments is ongoing.	No entanto, a investigação de novas tratamentos está a decorrer.	1						
53	The earlier PH is treated, the easier it is to control.	O pH for tratada, mais fácil se tornará para controlo .	1	2		3			
54	Treatments include medicines, procedures, and other therapies.	Tratamentos incluem medicação, procedimentos, entre outras; terapias.			2				
55	These treatments can relieve PH symptoms and slow the progress of the disease.	Estes tratamentos poderão aliviar os sintomas e pH atrasar a evolução da doença.	2	2					
56	Lifestyle changes also can help control symptoms.	Estilo de vida também pode ajudar a controlar os sintomas.				1			
57	Lifestyle changes	Estilo de vida							

58	- Monitor your weight: weigh yourself at the same time each day and record your weight in a diary or on a calendar.	- vigiar peso a: peso si na mesma hora do dia e registar peso a uma agenda ou a um calendário.	2			1		1	
59	- Lose weight if you are overweight (follow your doctor's recommendations).	- perder peso, se tiver excesso de peso, seguir as recomendações do seu médico.			2			1	
60	- Have annual health check-ups with your primary care physician.	- possuem anual saúde check-ups com o seu médico de família.	2				2	1	
61	- Activity guidelines: restrict lifting, pushing, or shoving to less than 20 pounds, since these activities increase the pressure in your arteries and lungs; walking is important to keep your muscles strong and improve circulation.	- actividade linhas orientadoras: restrição de se levantar, tanto ou shoving de menos de 20 kg, já que estas práticas aumento da Pressão nas artérias e pulmões. Caminhar é importante e manter os músculos fortes e otimizar a circulação.	2			1	3	1	
62	However, more strenuous forms of aerobic exercise should not be done regularly since your lungs may not be able to keep up with the oxygen demands placed on your body during these activities.	No entanto, mais árduas formas de aeróbica exercícios não deverão ser produzidos regularmente desde os seus pulmões, poderá não conseguir Mantanha-se a par das necessidades de oxigénio colocado no seu organismo, durante essas atividades.	3		1		2	2	
63	Any activities that cause shortness of breath, dizziness, or chest pain should be stopped immediately.	Todas as actividades que provocam falta de ar, vertigem ou dor torácica deve ser terminada imediatamente.	1						
64	Contact your doctor if any of these symptoms occur.	Contactar o seu médico se algum destes sintomas.				3			
65	- Do not smoke or chew tobacco.	- não fume ou mastigue tabaco.						1	
66	- Avoid or reduce alcohol intake.	- Evite ou Reduza o consumo de álcool.						1	
67	- Avoid pregnancy; use contraceptives to prevent pregnancy.	- evitar a gravidez; utilização de contraceptivos para evitar a gravidez.					1	1	1
68	- Maintain close contact with your health care providers.	- manter um contacto próximo com os profissionais de saúde.						1	1
69	When to call your doctor?	Quando contactar o seu médico?							
70	- Weight gain — 2 pounds in one day or 5 pounds in one week	- aumento de peso são alguns deles — 1 kg num dia ou 2 kg de seguimento				2	3	1	
71	- Swelling in your ankles, legs, or abdomen that has become worse	- se a tornozelos, pernas ou Abdomen, tornou-se pior			1	3		1	
72	- Shortness of breath that has become worse, especially if you wake up short of breath	- falta de ar que foram pior, especialmente se tiver dificuldades em acordar falta de ar				1	3	1	
73	- Extreme fatigue or decreased tolerance for activity	- fadiga extrema ou tolerância diminuída à actividade						1	

74	- A respiratory infection or a cough that has become worse	- uma infecção respiratória ou tosse, tornou-se pior	1		1	1		1	
75	- Fast heart rate (around 120 beats per minute)	Rápidas - frequência cardíaca (cerca de 120 batimentos por minuto)	2		1				
76	- Episodes of chest pain or discomfort with exertion that are not relieved with rest	- episódios de dor torácica nem desconforto com esforço, aliviado com repouso					2	1	
77	- Difficulty breathing with regular activities or at rest	- dificuldade para respirar com actividades habituais ou em repouso						1	
78	- Restlessness, confusion	- inquietação, confusão						1	
79	- Constant dizziness or lightheadedness	- constante tontura tonturas ou	2	1				1	
80	- Nausea or poor appetite	- náuseas, falta de apetite						1	
81	When to go to the emergency department?	Quando dirigir-se às Urgências?							
82	- Hickman catheter complications with intravenous Flolan, such as infection, catheter displacement/leakage of solution or blood, IV pump malfunction	- cateter de Hickman complicações com intravenosa flolan, tais como infecções, cateter displacement/perda de solução ou sangue, IV bomba tomar	3				3	1	
83	- New chest pain or discomfort that is severe, unexpected, and is accompanied by shortness of breath, sweating, nausea, or weakness	- nova dor torácica nem desconforto se for muito severa, inesperado, e é acompanhada de falta de ar, suores, náuseas ou fraqueza	2				3	1	
84	- Fast, sustained heart rate (more than 120 to 150 beats per minute) — especially if you are short of breath	- rapidez sustentada frequência cardíaca (mais de 120 a 150 batimentos por minuto) — sobretudo que falta de ar	3		1	3	2		
85	- Shortness of breath that is NOT relieved by rest	- falta de ar, Sem alívio em repouso						2	
86	- Sudden weakness or paralysis in your arms or legs	- fraqueza ou Paralisia nos braços ou pernas				2		1	
87	- Sudden onset of severe headache	- início súbito de dor de cabeça muito intensa						1	
88	- Fainting spells with loss of consciousness	- spells com perda da consciência desmaio	2				3	1	
89	Preeclampsia and Eclampsia	Pré-eclâmpsia e eclâmpsia							
90	Preeclampsia (toxemia) and eclampsia are conditions that can develop during week 20 or later of your pregnancy.	Pré-eclâmpsia (toxemia) e eclâmpsia são condições que desenvolvem durante 20 ou mais semanas de gravidez.					3		
91	With preeclampsia, you have high blood pressure and protein in your urine.	Com pré-eclâmpsia, Sofre de Hipertensão e proteínas na sua urina.			1		2	1	
92	Preeclampsia progresses to eclampsia if you have a seizure.	Pré-eclâmpsia progride para eclâmpsia se tiver uma convulsão.							
93	This does not include women with a known seizure disorder.	Isto não inclui mulheres com conhecidas epilepsia.	2						
94	These conditions can create mild to life-threatening health problems for you and your unborn baby.	Estas condições podem criar leve para que coloquem a vida em risco problemas de saúde do seu unborn bebé.	3				3		

95	You will need tests 1 to 2 times a week to check your condition.	De testes de 1 a 2 vezes por semana, para avaliar o teu estado de saúde.	2			3			
96	Tests include blood pressure checks, urine and blood tests, and fetal monitoring.	Testes incluir Pressão arterial verifica, urina, análises, monitorização fetal.	2					1	
97	Rest:	Repouso:							
98	- Your obstetrician may tell you to rest more often if you have mild symptoms of preeclampsia.	- o obstetra pode dizer que Durma para descansar com mais frequência se apresentar sintomas ligeiros de pré-eclâmpsia.					2	1	
99	You may need total bedrest if you have more severe symptoms.	Pode precisar de total de repouso na cama mais se apresentar sintomas severos.	2						
100	Try to lie on your left side.	Tente Deite-se de barriga lado esquerdo.	1			1		1	

iii. “Instruções de Alta” – Avaliação da tradução do Moses for Mere Mortals (EMEA)

	ORIGINAL_EN	MMM_PT (EMEA)	Gr a	Ter	Po	Om	Tra	Mai	Coe
1	Chronic hypertension	Crónica hipertensão	1						
2	Chronic hypertension is a long-term condition in which your blood pressure (BP) is higher than normal.	Crónica a hipertensão é uma doença de longa duração em que a tensão arterial (PA) é superior à normal.	1						1
3	Your BP is the force of your blood moving against the walls of your arteries.	Sua tensão arterial está a força da movimentação contra as paredes das suas artérias.				2	2		1
4	Normal blood pressure is less than 120/80.	Sem tensão arterial for inferior a 120/80.				2	3		1
5	Prehypertension is BP between 120/80 and 139/89.	Prehypertension é entre PA 120/80 e 151/93.	1				2		
6	High blood pressure is 140/90 or higher.	Pressão arterial elevada é de 140/90 ou superiores.	1			1			
7	Medicines that may be ordered for you	Os medicamentos que podem ser prescritos para si							
8	- Blood pressure medicine is given to lower your blood pressure.	- pressão sanguínea medicamento é administrado para baixar a sua pressão sanguínea.	1					1	
9	A controlled blood pressure helps protect your organs, such as your heart, lungs, brain, and kidneys.	Controlada (tensão) ajuda a proteger os órgãos, como o coração, pulmão, cérebro e rins.	1		1				1
10	You may need more than 1 type of blood pressure medicine.	Poderá ter mais do que um tipo de pressão arterial.				3	3		
11	Take your blood pressure medicine exactly as directed.	Tome a sua tensão arterial medicamento exactamente como prescrito.	1						1
12	- Diuretics help decrease extra fluid that collects in your body.	Deverá ajudar a diminuir de fluido que faz no seu organismo.				3	3		
13	This will help lower your BP.	Isto ajuda a baixar a tensão arterial desejada.				3			
14	You may urinate more often while you take this medicine.	Pode urinar mais frequentemente enquanto toma este medicamento.	1						
15	- Take your medicine as directed.	- Tome a medicação conforme indicado.							
16	Call your healthcare provider if you think your medicine is not helping or if you have side effects.	Contacte o seu profissional de saúde se pensa que o medicamento não se ajudar, ou se tiver efeitos secundários.				2	1		
17	Tell him if you are allergic to any medicine.	Informe ele se tem alergia (hipersensibilidade) a qualquer medicamento.	2						
18	Keep a list of the medicines, vitamins, and herbs you take.	Mantenha uma lista das vitaminas, produtos de ervanária ou tenha tomado.				2	1		
19	Include the amounts, and when and why you take them.	Incluem os montantes e, quando e por si tomá-los.	3				3		

20	Bring the list or the pill bottles to follow-up visits.	Término da lista e as embalagens dos seus medicamentos para as consultas de seguimento.					3		
21	Carry your medicine list with you in case of an emergency.	Traga a sua medicação consigo uma lista em caso de emergência.	2						
22	Pulmonary hypertension	Hipertensão pulmonar							
23	Pulmonary hypertension, or PH, is increased pressure in the pulmonary arteries.	Hipertensão pulmonar, ou de pH, aumento da pressão da artéria pulmonar.		2		1	1		
24	These arteries carry blood from your heart to your lungs to pick up oxygen.	Estas artérias transportam o sangue do coração aos pulmões pegar oxigénio.	1			2	1		
25	PH causes symptoms such as shortness of breath during routine activity (for example, climbing two flights of stairs), tiredness, chest pain, and a racing heartbeat.	PH causa sintomas tais como falta de ar durante actividades de rotina (por exemplo, dois subir escadas de voos), fadiga, dor torácica e o ritmo cardíaco acelerado.	1				2		
26	As the condition worsens, its symptoms may limit all physical activity.	Como a doença piorar, os seus sintomas podem limitar as actividades físicas.				1	2		
27	To understand PH, it helps to understand how your heart and lungs work.	Perceber pH, ajuda a perceber como o coração e os pulmões.		2		3			
28	Your heart has two sides, separated by an inner wall called the septum.	O coração tem dois lados, separados por parede do septo.				3			
29	Each side of your heart has an upper and lower chamber.	Cada lado do seu superior e inferior do tanque.				3	1		
30	The lower right chamber of your heart, the right ventricle, pumps blood to your pulmonary arteries.	Canto inferior direito da câmara de coração, do ventrículo direito, aspiradores de sangue ao artérias pulmonares.	3				3		
31	The blood then travels to your lungs, where it picks up oxygen.	Sanguíneos de Viagens aos pulmões, aonde picks de oxigénio.	3				3		
32	The upper left chamber of your heart, the left atrium, receives the oxygen-rich blood from your lungs.	Canto superior esquerdo da câmara de coração, do Átrio esquerdo, recepção do coração de fazer chegar sangue rico dos pulmões.	3			3	3	1	1
33	The blood is then pumped into the lower left chamber of your heart, the left ventricle.	O sangue é bombeado para dentro da câmara inferior esquerdo do coração, do ventrículo esquerdo.	1				1		
34	From the left ventricle, the blood is pumped to the rest of your body through an artery called the aorta.	Do ventrículo esquerdo, O sangue é bombeado para o restante organismo através de uma artéria aorta, chamada.	1		2				
35	PH begins with inflammation and changes in the cells that line your pulmonary arteries.	PH iniciado com inflamação e alterações nas células que revestem as artérias pulmonares.	1				2		
36	Other factors also can affect the pulmonary arteries and cause PH.	Outros Fatores poderão igualmente afectar as artérias pulmonares e causar o pH.		2				1	
37	For example, the condition may develop if:	Por exemplo, a condição pode ocorrer se:					1		1

38	- The walls of the arteries tighten.	- as paredes das artérias apertar.	1				1	1	
39	- The walls of the arteries are stiff at birth or become stiff from an overgrowth of cells.	- as paredes das artérias rígidos à nascença ou ficar rígida da proliferação de células.	1			2			
40	- Blood clots form in the arteries.	- formar coágulos sanguíneos nas artérias.	1					1	
41	- These changes make it hard for your heart to push blood through your pulmonary arteries and into your lungs.	- estas modificações que este se torne para o coração). A pressão sanguínea pulmonar através das artérias e para os seus pulmões.	3		3	3	3	1	
42	As a result, the pressure in your arteries rises.	Como resultado, a pressão nos seus artérias aumento	1		1				
43	Also, because your heart is working harder than normal, your right ventricle becomes strained and weak.	Também porque o seu coração está a funcionar mais do que o normal, o seu direito ventrículo fica tendão fraco.	1			1	3		
44	Your heart may become so weak that it can't pump enough blood to your lungs.	O seu coração pode tornar-se debilitado, para que este não consegue bombear sangue suficiente para os pulmões.				1	2		
45	This causes heart failure.	A insuficiência cardíaca.				3			
46	Heart failure is the most common cause of death in people who have PH.	Insuficiência cardíaca é a causa de morte mais frequente em doentes com Te.		3					
47	PH is divided into five groups based on its causes and treatment options.	PH está dividido em cinco grupos com base nas suas causas e opções terapêuticas.	1				2		
48	In all groups, the average pressure in the pulmonary arteries is 25 mmHg or higher.	, 4%, as médias na pressão da artéria pulmonar é de 25 mm Hg ou superior.			2	3	2	1	
49	The pressure in normal pulmonary arteries is 8–20 mmHg at rest.	Da pressão da artéria pulmonar normal é de 8- 20 mm/Hg em repouso.	2				1		
50	(The mmHg is millimeters of mercury—the units used to measure blood pressure.)	(mmHg é millimeters de mercúrio — as unidades utilizadas para medir a pressão arterial.)					1		
51	PH has no cure.	PH não cura.	1			3	2		
52	However, research for new treatments is ongoing.	Contudo, investigação para novos tratamentos está a decorrer.	1						
53	The earlier PH is treated, the easier it is to control.	O pH é tratada, mais fácil se tornará para o controlo.		2		3	1		
54	Treatments include medicines, procedures, and other therapies.	Tratamentos incluem medicamentos, procedimentos, entre outros tratamentos.							
55	These treatments can relieve PH symptoms and slow the progress of the disease.	O pH pode aliviar os sintomas e retardar a evolução da doença.		2		3			
56	Lifestyle changes also can help control symptoms.	O estilo de vida Também poderá ajudar a controlar os sintomas.			2				
57	Lifestyle changes	O estilo de vida							

58	- Monitor your weight: weigh yourself at the same time each day and record your weight in a diary or on a calendar.	- vigiar o seu peso: peso por si , à mesma hora todos os dias e registar os seus num diário ou num calendário.	1			2	2	1	
59	- Lose weight if you are overweight (follow your doctor's recommendations).	- perder peso, se tiver excesso de peso, siga as instruções do seu médico.).	1		2			1	
60	- Have annual health check-ups with your primary care physician.	- já anual de rotina com o seu médico de família.				3			
61	- Activity guidelines: restrict lifting, pushing, or shoving to less than 20 pounds, since these activities increase the pressure in your arteries and lungs; walking is important to keep your muscles strong and improve circulation.	- atividade e documentos restringe o levantamento de pesos , empurrando ou shoving para menos de 20 kg, uma vez que estas actividades aumento da pressão nas suas artérias pulmonares; caminhar é importante manter os músculos fortes e a circulação melhora .	3			3	3		
62	However, more strenuous forms of aerobic exercise should not be done regularly since your lungs may not be able to keep up with the oxygen demands placed on your body during these activities.	No entanto, mais intenso formas de exercícios de aeróbica não devem ser efectuadas regularmente pois seus pulmões de acompanhar as exigências de oxigénio colocado no seu corpo durante estas actividades;	2			3			
63	Any activities that cause shortness of breath, dizziness, or chest pain should be stopped immediately.	Quaisquer actividades que cause falta de ar, vertigem ou dor no peito , devem ser imediatamente suspensas.	1						1
64	Contact your doctor if any of these symptoms occur.	Contactar o médico caso surja algum destes sintomas.							
65	- Do not smoke or chew tobacco.	- não fume ou mastigue tabaco.						1	
66	- Avoid or reduce alcohol intake.	- Evite ou reduza o consumo de álcool.							
67	- Avoid pregnancy; use contraceptives to prevent pregnancy.	- evitar a gravidez. O uso de contraceptivos para evitar a gravidez.					2	1	1
68	- Maintain close contact with your health care providers.	- manter contacto próximo com os seus profissionais de saúde.						1	1
69	When to call your doctor?	Quando contactar o médico?							
70	- Weight gain — 2 pounds in one day or 5 pounds in one week	- aumento de peso - 1 kg num dia ou de 2 kg de			1	3		1	
71	- Swelling in your ankles, legs, or abdomen that has become worse	- inchaço dos tornozelos, pernas, abdómen ou que tenha piorado				2		1	
72	- Shortness of breath that has become worse, especialmente se you wake up short of breath	- falta de ar que tenha piorado, acordar com falta de ar				2		1	
73	- Extreme fatigue or decreased tolerance for activity	- intensa fadiga ou diminuição da tolerância da actividade	2					1	

74	- A respiratory infection or a cough that has become worse	- uma infecção respiratória ou tosse que tenha piorado						1	
75	- Fast heart rate (around 120 beats per minute)	- frequência cardíaca aumentada (cerca de 120 bpm)						1	
76	- Episodes of chest pain or discomfort with exertion that are not relieved with rest	- episódios de dor no peito ou desconforto abdominal com esforço não melhora com repouso				1		1	1
77	- Difficulty breathing with regular activities or at rest	- dificuldade em respirar com actividades habituais ou em repouso						1	
78	- Restlessness, confusion	- inquietação, confusão						1	
79	- Constant dizziness or lightheadedness	- constante de tonturas ou atordoamento	2					1	
80	- Nausea or poor appetite	- náuseas, estômago inchado					3	1	
81	When to go to the emergency department?	Quando deverá dirigir-se às urgências?							
82	- Hickman catheter complications with intravenous Flolan, such as infection, catheter displacement/leakage of solution or blood, IV pump malfunction	- cateter de Hickman complicações com intravenosa flolan, tais como infecção, cateter deslocação/ou perdas de sangue, soros mau funcionamento	3		1	3		1	
83	- New chest pain or discomfort that is severe, unexpected, and is accompanied by shortness of breath, sweating, nausea, or weakness	- nova dor ou desconforto no peito que se graves e inesperadas, sendo acompanhada pela falta de ar, suores, náuseas, fraqueza ou	2			2		1	1
84	- Fast, sustained heart rate (more than 120 to 150 beats per minute) — especially if you are short of breath	- rápida e sustentada da frequência cardíaca (mais de 120 a 100 batimentos por minuto) — sobretudo se tem dificuldade em respirar	2					1	
85	- Shortness of breath that is NOT relieved by rest	- falta de ar que não for aliviada pelo repouso	1					2	
86	- Sudden weakness or paralysis in your arms or legs	- fraqueza ou paralisia nos braços ou pernas				3		1	
87	- Sudden onset of severe headache	- início súbito de cefaleias graves						1	
88	- Fainting spells with loss of consciousness	- desmaios, perda de consciência momentânea com	3					1	
89	Preeclampsia and Eclampsia	E pré-eclâmpsia eclâmpsia				2			
90	Preeclampsia (toxemia) and eclampsia are conditions that can develop during week 20 or later of your pregnancy.	Pré-eclâmpsia (toxemia) e eclâmpsia são situações que podem surgir durante 20 ou mais semanas de gestação.					3		1
91	With preeclampsia, you have high blood pressure and protein in your urine.	Com pré-eclâmpsia tem tensão arterial elevada e de proteínas na urina.				2			
92	Preeclampsia progresses to eclampsia if you have a seizure.	Pré-eclâmpsia progride para eclâmpsia se tiver uma convulsão.							
93	This does not include women with a known seizure disorder.	Isto não inclui mulheres com epilepsia.							

94	These conditions can create mild to life-threatening health problems for you and your unborn baby.	Estas condições podem ser criados ligeira a potencialmente fatal problemas de saúde para si e para o seu filho.	2				3		1
95	You will need tests 1 to 2 times a week to check your condition.	Necessitará de 1 a 2 vezes por semana, para controlar a sua doença.				3			
96	Tests include blood pressure checks, urine and blood tests, and fetal monitoring.	Incluem testes da tensão arterial permite, urina e ao sangue e monitorização fetal.	3			3			
97	Rest:	Repouso:							
98	- Your obstetrician may tell you to rest more often if you have mild symptoms of preeclampsia.	- o (a) médico obstetra pode dizer- lhe para descansar com maior frequência se tiver sintomas ligeiros de pré-eclâmpsia.							
99	You may need total bedrest if you have more severe symptoms.	Pode necessitar de um total de repouso na cama se tiver sintomas mais graves.	1						
100	Try to lie on your left side.	Tente Deite-se lado esquerdo.	1			2		1	2

iv. “Instruções de Alta” – Avaliação da tradução do Moses for Mere Mortals (Europarl)

	ORIGINAL_EN	MMM_PT (Europarl)	Gr a	Ter	Po	Om	Tra	Mai	Coe
1	Chronic hypertension	Crónica hipertensão	1						
2	Chronic hypertension is a long-term condition in which your blood pressure (BP) is higher than normal.	Crónica a hipertensão é uma doença na qual a pressão arterial (PA) é superior à normal.	1						
3	Your BP is the force of your blood moving against the walls of your arteries.	A BP está a força do seu sangue a de ir contra as paredes das suas artérias.	2				2		2
4	Normal blood pressure is less than 120/80.	Normais de pressão arterial inferior a 120/80.	2			2			
5	Prehypertension is BP between 120/80 and 139/89.	Prehypertension é a BP entre 120/80 e 151/89.					2		2
6	High blood pressure is 140/90 or higher.	Pressão arterial elevada é de 140/90 mg ou superiores.	1						
7	Medicines that may be ordered for you	Os medicamentos que podem ser prescritos para vos	1						
8	- Blood pressure medicine is given to lower your blood pressure.	- hipertensão Medicina é dada para baixar a sua pressão sanguínea.	2				3	1	
9	A controlled blood pressure helps protect your organs, such as your heart, lungs, brain, and kidneys.	Uma tensão arterial controlada, protegemos os seus órgãos, tais como problemas no coração, pulmão, cérebro e rins.	2			3			
10	You may need more than 1 type of blood pressure medicine.	Pode ser necessário mais do que um tipo de pressão arterial.				3			
11	Take your blood pressure medicine exactly as directed.	Tomar a medicação exatamente conforme as instruções.				2			
12	- Diuretics help decrease extra fluid that collects in your body.	- diuréticos ajudar a diminuir de fluido em excesso que recebe no seu corpo.	2				2	1	
13	This will help lower your BP.	Isto ajuda a baixar a pressão arterial.							
14	You may urinate more often while you take this medicine.	Pode urinar mais frequentemente enquanto toma este medicamento.	1				1		
15	- Take your medicine as directed.	- tome a medicação conforme as instruções.						1	
16	Call your healthcare provider if you think your medicine is not helping or if you have side effects.	Contacte o seu profissional de saúde se pensa que o medicamento não está a resultar e se tiver efeitos secundários.					1		
17	Tell him if you are allergic to any medicine.	Dizer-lhe se tem alergia (hipersensibilidade) a qualquer medicamento.							
18	Keep a list of the medicines, vitamins, and herbs you take.	Mantenha uma lista das vitaminas, produtos de ervanária ou tomar.	2			2			

19	Include the amounts, and when and why you take them.	Incluir os montantes, quando e por que motivo os toma.					2		1
20	Bring the list or the pill bottles to follow-up visits.	Trazer a lista e as embalagens dos seus medicamentos para as consultas de seguimento de uma pílula.				3			1
21	Carry your medicine list with you in case of an emergency.	Leve o medicamento consigo enumerar, em caso de incidente.				3	3		
22	Pulmonary hypertension	Pulmonary hypertension					3		
23	Pulmonary hypertension, or PH, is increased pressure in the pulmonary arteries.	Hipertensão pulmonar, ou de pH, aumenta a pressão nas artérias pulmonares.		2		3			
24	These arteries carry blood from your heart to your lungs to pick up oxygen.	Essas artérias que transportam o sangue do coração aos pulmões, de pegar oxigénio.				3	2		
25	PH causes symptoms such as shortness of breath during routine activity (for example, climbing two flights of stairs), tiredness, chest pain, and a racing heartbeat.	PH causa sintomas tais como falta de ar durante actividades usuais (por exemplo, construir dois voos de escadas), cansaço, dor no peito e uma aceleração do ritmo cardíaco.	1				3		
26	As the condition worsens, its symptoms may limit all physical activity.	Como a situação se deteriore. Os seus sintomas podem limitar toda a actividade física.	1		2		3		
27	To understand PH, it helps to understand how your heart and lungs work.	Compreender pH, ajuda a perceber como o coração e os pulmões.		2		3			
28	Your heart has two sides, separated by an inner wall called the septum.	Coração tem duas vertentes, separados por uma parede, o septo.				1	2		
29	Each side of your heart has an upper and lower chamber.	Cada lado do seu coração é uma máxima e sala.					3		
30	The lower right chamber of your heart, the right ventricle, pumps blood to your pulmonary arteries.	Na zona inferior direita do hemiciclo, do ventrículo direito, bombas de sangue para as artérias pulmonares.	2				3		
31	The blood then travels to your lungs, where it picks up oxygen.	Sanguínea, pois viaja para os pulmões, onde se regista de oxigénio.					3		
32	The upper left chamber of your heart, the left atrium, receives the oxygen-rich blood from your lungs.	Parte superior esquerda do hemiciclo, no Átrio esquerdo, recebe oxygen-rich sangue dos pulmões.	1			1	3		1
33	The blood is then pumped into the lower left chamber of your heart, the left ventricle.	O sangue é bombeado para a parte inferior esquerda Câmara do ventrículo esquerdo.	2		1		2	1	
34	From the left ventricle, the blood is pumped to the rest of your body through an artery called the aorta.	Do ventrículo esquerdo, O sangue é bombeado para o restante organismo através de uma artéria aorta chamada.	1		2				
35	PH begins with inflammation and changes in the cells that line your pulmonary arteries.	PH começa pela inflamação e alterações nas células que revestem as artérias pulmonares.	1				2		

36	Other factors also can affect the pulmonary arteries and cause PH.	Outros factores que também podem afectar as artérias pulmonares e provocar o pH.		2					
37	For example, the condition may develop if:	Por exemplo, a situação pode ocorrer se:					3		
38	- The walls of the arteries tighten.	- parede de uma artéria apertar.					3	1	
39	- The walls of the arteries are stiff at birth or become stiff from an overgrowth of cells.	- parede de uma artéria rígidos à nascença ou se tornam rígido de uma proliferação de células.	2			2	2		
40	- Blood clots form in the arteries.	- formar coágulos sanguíneos nas artérias.	1					1	
41	- These changes make it hard for your heart to push blood through your pulmonary arteries and into your lungs.	- estas alterações tornam difícil ao coração de sangue nas suas artérias pulmonares e para os seus pulmões.				3	2	1	
42	As a result, the pressure in your arteries rises.	Como resultado, a pressão nas suas artérias.				3			
43	Also, because your heart is working harder than normal, your right ventricle becomes strained and weak.	Também, porque o seu coração está a funcionar de uma forma mais do que o normal, o seu ventrículo direito for condicionado e fracos.	1			3	3		
44	Your heart may become so weak that it can't pump enough blood to your lungs.	O seu coração pode tornar-se tão fracas que não consegue bombear sangue suficiente para os pulmões.	1						
45	This causes heart failure.	Isto causa problemas cardiovasculares.		1					
46	Heart failure is the most common cause of death in people who have PH.	Insuficiência cardíaca é a causa de morte mais comum nas pessoas que apresentem LLA pH.		2		2			1
47	PH is divided into five groups based on its causes and treatment options.	PH está dividido em cinco grupos com base na sua origem e as alternativas terapêuticas.	1				2		
48	In all groups, the average pressure in the pulmonary arteries is 25 mmHg or higher.	Em todos os grupos, a média na pressão da artéria pulmonar de 25 mm Hg ou superior.	2			2			
49	The pressure in normal pulmonary arteries is 8–20 mmHg at rest.	A pressão da artéria pulmonar normal é de 8- 20 mm/Hg em repouso.	1						
50	(The mmHg is millimeters of mercury—the units used to measure blood pressure.)	(mmHg é millimeters de mercúrio -, as unidades utilizadas para medir a pressão arterial.)			1		1		
51	PH has no cure.	PH não tem remédio.	1				2		
52	However, research for new treatments is ongoing.	Contudo, a investigação de novos tratamentos.				3			
53	The earlier PH is treated, the easier it is to control.	A um pH mais cedo for tratada, mais fácil será controlar.		2			2		
54	Treatments include medicines, procedures, and other therapies.	Tratamentos incluem medicamentos, procedimentos, entre outros tratamentos.					1		
55	These treatments can relieve PH symptoms and slow the progress of the disease.	Estes tratamentos pode aliviar os sintomas de pH e lenta a evolução da doença.	1	2			2		

56	Lifestyle changes also can help control symptoms.	As alterações ao estilo de vida que também pode ajudar a controlar os sintomas.	1			2			
57	Lifestyle changes	O estilo					3		
58	- Monitor your weight: weigh yourself at the same time each day and record your weight in a diary or on a calendar.	- vigiar o seu peso: peso Exa. sempre à mesma hora e no registo de peso numa agenda ou a um calendário.					3	1	
59	- Lose weight if you are overweight (follow your doctor's recommendations).	- perder peso, se tiver excesso de peso (siga as instruções do seu médico.).			1			1	
60	- Have annual health check-ups with your primary care physician.	- realização de exames médicos anuais com o seu médico de família.						1	1
61	- Activity guidelines: restrict lifting, pushing, or shoving to less than 20 pounds, since these activities increase the pressure in your arteries and lungs; walking is important to keep your muscles strong and improve circulation.	- directrizes relativas à actividade: restringir o levantamento, posse ou jogos para menos de 20 kg, uma vez que estas actividades de aumentar a pressão nas suas artérias pulmonares e caminhar é importante manter os músculos fortes e o melhoramento da circulação.					3	1	
62	However, more strenuous forms of aerobic exercise should not be done regularly since your lungs may not be able to keep up with the oxygen demands placed on your body during these activities.	Entanto mais enérgicos tipos de exercícios de aeróbica não deve ser efectuada com regularidade desde que os seus pulmões pode não ser capaz de satisfazer as exigências de oxigénio no corpo durante estas actividades;	3				3		
63	Any activities that cause shortness of breath, dizziness, or chest pain should be stopped immediately.	Quaisquer actividades que causam dificuldade em respirar, tonturas, dor no peito e devem ser imediatamente suspensas.	1						
64	Contact your doctor if any of these symptoms occur.	Informe imediatamente o seu médico se tiver algum destes sintomas.							
65	- Do not smoke or chew tobacco.	- não fume ou mastigue tabaco.						1	
66	- Avoid or reduce alcohol intake.	- evite ou reduza o consumo de álcool.						1	
67	- Avoid pregnancy; use contraceptives to prevent pregnancy.	- evitar a gravidez. O uso de contraceptivos para evitar a gravidez.					2	1	1
68	- Maintain close contact with your health care providers.	- manter um estreito contacto com os seus profissionais de saúde.						1	1
69	When to call your doctor?	Quando contactar o médico?							
70	- Weight gain — 2 pounds in one day or 5 pounds in one week	- aumento de peso - 2 libras num dia ou 2 kg numa semana					1	1	2
71	- Swelling in your ankles, legs, or abdomen that has become worse	- inchaço dos tornozelos, pernas, abdómen ou que tenha piorado				2		1	

72	- Shortness of breath that has become worse, especially if you wake up short of breath	- falta de ar que se agravou, sobretudo se lhe deparam , com pouco fôlego	1					3		
73	- Extreme fatigue or decreased tolerance for activity	- intensa fadiga ou diminuição da tolerância à actividade	1						1	
74	- A respiratory infection or a cough that has become worse	- uma infecção respiratória ou tosse, que se agravou			1				1	
75	- Fast heart rate (around 120 beats per minute)	- aumento da frequência cardíaca (bpm). Cerca de 120	3						1	
76	- Episodes of chest pain or discomfort with exertion that are not relieved with rest	- episódios de dor no peito ou desconforto abdominal com esforço, sem alívio em repouso							1	
77	- Difficulty breathing with regular activities or at rest	- dificuldade em respirar com actividades habituais ou em repouso							1	
78	- Restlessness, confusion	- agitação, Estado confusional							1	
79	- Constant dizziness or lightheadedness	Constante de tonturas ou atordoamento	2		1					
80	- Nausea or poor appetite	- náuseas, perda de apetite							1	
81	When to go to the emergency department?	Quando deverá dirigir-se às urgências?								
82	- Hickman catheter complications with intravenous Flolan, such as infection, catheter displacement/leakage of solution or blood, IV pump malfunction	Cateter de Hickman - complicações com intravenosa flolan, tais como infecção, cateter deslocamento/ou perdas de sangue, soros avaria na bomba	3		1					
83	- New chest pain or discomfort that is severe, unexpected, and is accompanied by shortness of breath, sweating, nausea, or weakness	Novo - dor ou desconforto no peito que é grave, inesperado, e é acompanhada pela falta de ar, suores, náuseas, fraqueza ou	3							
84	- Fast, sustained heart rate (more than 120 to 150 beats per minute) — especially if you are short of breath	- rápida e sustentada da frequência cardíaca (mais de 120 a 100 batimentos por minuto), especialmente se tiver dificuldade em respirar	1						1	
85	- Shortness of breath that is NOT relieved by rest	- falta de ar que não for aliviada pelo repouso	1						1	
86	- Sudden weakness or paralysis in your arms or legs	- súbita, ou paralisia, fraqueza nos braços ou pernas	1		2				1	
87	- Sudden onset of severe headache	- início súbito de cefaleias graves							1	
88	- Fainting spells with loss of consciousness	- desmaios, perda de consciência momentânea com	2							
89	Preeclampsia and Eclampsia	E pré-eclâmpsia eclâmpsia	2							
90	Preeclampsia (toxemia) and eclampsia are conditions that can develop during week 20 or later of your pregnancy.	Pré-eclâmpsia () e toxemia eclâmpsia são condições que possam desenvolver-se durante 20 ou mais semanas de gestação.	2					3		
91	With preeclampsia, you have high blood pressure and protein in your urine.	De pré-eclâmpsia, tem tensão arterial elevada e de proteínas na urina.					2	2		
92	Preeclampsia progresses to eclampsia if you have a seizure.	Pré-eclâmpsia progride a eclampsia se tiver uma convulsão.	1							

93	This does not include women with a known seizure disorder.	Isto não inclui mulheres com epilepsia.							
94	These conditions can create mild to life-threatening health problems for you and your unborn baby.	Estas situações podem criar problemas de saúde de ligeira a potencialmente fatal para si e para o seu filho.	1						
95	You will need tests 1 to 2 times a week to check your condition.	Irá necessitar de análises ao sangue após 1 ou 2 dias por semana, para controlar a sua doença.					3		
96	Tests include blood pressure checks, urine and blood tests, and fetal monitoring.	Incluem testes de tensão controlo da urina e ao sangue e monitorização fetal.	3						
97	Rest:	Repouso:							
98	- Your obstetrician may tell you to rest more often if you have mild symptoms of preeclampsia.	- o obstetra pode dizer- lhe para ficar com mais frequência se tiver sintomas ligeiros de pré-eclâmpsia.					3		
99	You may need total bedrest if you have more severe symptoms.	Pode necessitar de um total de repouso na cama. Se mais graves, os sintomas.	3		3				
100	Try to lie on your left side.	Tentar Deite-se de esquerda.	1				1	1	1

v. “Instruções de Alta” – Avaliação da tradução do Google Tradutor

	ORIGINAL_EN	PT (Google)	Gr a	Ter	Po	Om	Tra	Mai	Coe
1	Chronic hypertension	hipertensão crônica						1	
2	Chronic hypertension is a long-term condition in which your blood pressure (BP) is higher than normal.	Hipertensão crônica é uma condição de longo prazo em que a sua pressão arterial (PA) é maior do que o normal.							
3	Your BP is the force of your blood moving against the walls of your arteries.	O BP é a força de seu sangue movendo-se contra as paredes de suas artérias.							2
4	Normal blood pressure is less than 120/80.	Pressão arterial normal é inferior a 120/80.							
5	Prehypertension is BP between 120/80 and 139/89.	Pré-hipertensão é BP entre 120/80 e 139/89.							2
6	High blood pressure is 140/90 or higher.	A hipertensão arterial é 140/90 ou superior.							
7	Medicines that may be ordered for you	Os medicamentos que podem ser encomendados para você							
8	- Blood pressure medicine is given to lower your blood pressure.	- Pressão arterial medicina é dado para baixar a pressão arterial.					3		
9	A controlled blood pressure helps protect your organs, such as your heart, lungs, brain, and kidneys.	A pressão arterial controlada ajuda a proteger seus órgãos, como o coração, pulmões, cérebro e rins.							
10	You may need more than 1 type of blood pressure medicine.	Você pode precisar de mais de um tipo de medicamento para pressão alta.							
11	Take your blood pressure medicine exactly as directed.	Tome o seu medicamento a pressão arterial exatamente como dirigido.					1		
12	- Diuretics help decrease extra fluid that collects in your body.	- Diminuição Diuréticos ajuda fluido extra que se acumula em seu corpo.	2				3		
13	This will help lower your BP.	Isso vai ajudar a diminuir o BP.					2		
14	You may urinate more often while you take this medicine.	Você pode urinar mais frequentemente enquanto tomar este medicamento.							
15	- Take your medicine as directed.	- Tome o seu medicamento como prescrito.							
16	Call your healthcare provider if you think your medicine is not helping or if you have side effects.	Chame seu médico se você acha que o seu medicamento não está ajudando ou se tiver efeitos secundários.							
17	Tell him if you are allergic to any medicine.	Diga-lhe que, se você é alérgico a qualquer medicamento.			1		1		
18	Keep a list of the medicines, vitamins, and herbs you take.	Mantenha uma lista de medicamentos, vitaminas e ervas que você toma.							
19	Include the amounts, and when and why you take them.	Inclui os valores, e quando e por que você levá-los.	1				3		

20	Bring the list or the pill bottles to follow-up visits.	Trazer a lista ou os frascos de comprimidos para acompanhar as visitas.	1					3		
21	Carry your medicine list with you in case of an emergency.	Leve sua lista de medicamentos com você em caso de uma emergência.								
22	Pulmonary hypertension	A hipertensão pulmonar								
23	Pulmonary hypertension, or PH, is increased pressure in the pulmonary arteries.	A hipertensão pulmonar, ou PH, é o aumento da pressão nas artérias pulmonares.						2		
24	These arteries carry blood from your heart to your lungs to pick up oxygen.	Estas artérias levam o sangue do seu coração para os pulmões para pegar oxigênio.								
25	PH causes symptoms such as shortness of breath during routine activity (for example, climbing two flights of stairs), tiredness, chest pain, and a racing heartbeat.	PH provoca sintomas como falta de ar durante a atividade de rotina (por exemplo, subir dois lances de escada), cansaço, dor no peito e palpitações cardíacas.	1					2		
26	As the condition worsens, its symptoms may limit all physical activity.	Como o quadro se agrava, seus sintomas podem limitar toda a atividade física.						2		
27	To understand PH, it helps to understand how your heart and lungs work.	Para entender o PH, que ajuda a entender como o seu coração e os pulmões de trabalho.					1	2		
28	Your heart has two sides, separated by an inner wall called the septum.	Seu coração tem dois lados, separados por uma parede interna chamada septo.								
29	Each side of your heart has an upper and lower chamber.	Cada lado do seu coração tem uma câmara superior e inferior.								
30	The lower right chamber of your heart, the right ventricle, pumps blood to your pulmonary arteries.	A câmara inferior direito do seu coração, o ventrículo direito, bombeia sangue para as artérias pulmonares.								
31	The blood then travels to your lungs, where it picks up oxygen.	O sangue, em seguida, viaja para os pulmões, onde ele capta oxigênio.								
32	The upper left chamber of your heart, the left atrium, receives the oxygen-rich blood from your lungs.	A câmara superior esquerda do coração, o átrio esquerdo, recebe o sangue rico em oxigênio dos pulmões.								
33	The blood is then pumped into the lower left chamber of your heart, the left ventricle.	O sangue é bombeado para a câmara inferior esquerdo do seu coração, o ventrículo esquerdo.	1							
34	From the left ventricle, the blood is pumped to the rest of your body through an artery called the aorta.	A partir do ventrículo esquerdo, o sangue é bombeado para o resto do seu corpo, através de uma artéria aorta chamado.	1							
35	PH begins with inflammation and changes in the cells that line your pulmonary arteries.	PH começa com a inflamação e alterações nas células que revestem as artérias pulmonares.	1					2		
36	Other factors also can affect the pulmonary arteries and cause PH.	Outros fatores também podem afetar as artérias pulmonares e causar PH.						2		

37	For example, the condition may develop if :	Por exemplo, a condição pode desenvolver-se:				2			
38	- The walls of the arteries tighten.	- As paredes das artérias apertar .	1						
39	- The walls of the arteries are stiff at birth or become stiff from an overgrowth of cells.	- As paredes das artérias são rígidas no nascimento ou tornar-se duro de um crescimento excessivo de células.	1				2		
40	- Blood clots form in the arteries.	- Formar coágulos de sangue nas artérias.	1						
41	- These changes make it hard for your heart to push blood through your pulmonary arteries and into your lungs.	- Estas alterações tornam difícil para o seu coração para empurrar o sangue através das artérias pulmonares e em seus pulmões.	2						
42	As a result, the pressure in your arteries rises.	Como resultado, a pressão nas artérias aumenta.							
43	Also, because your heart is working harder than normal, your right ventricle becomes strained and weak.	Também , porque o seu coração está trabalhando mais duro do que o normal, o ventrículo direito torna-se tensa e fraca .	1				1		
44	Your heart may become so weak that it can't pump enough blood to your lungs.	Seu coração pode tornar-se tão fraco que não consegue bombear sangue suficiente para os pulmões.							
45	This causes heart failure.	Isso faz com que a insuficiência cardíaca.					3		
46	Heart failure is the most common cause of death in people who have PH.	A insuficiência cardíaca é a causa mais comum de morte em pessoas que têm PH .					2		
47	PH is divided into five groups based on its causes and treatment options.	PH é dividido em cinco grupos de acordo com suas causas e opções de tratamento.	1				2		
48	In all groups, the average pressure in the pulmonary arteries is 25 mmHg or higher.	Em todos os grupos, a pressão média nas artérias pulmonares é de 25 mmHg ou superior.							
49	The pressure in normal pulmonary arteries is 8–20 mmHg at rest.	A pressão nas artérias pulmonares normais é 8-20 mmHg em repouso.	2						
50	(The mmHg is millimeters of mercury—the units used to measure blood pressure.)	(O mmHg é milímetros de mercúrio—as unidades usadas para medir a pressão arterial).							
51	PH has no cure.	PH não tem cura.	1				2		
52	However, research for new treatments is ongoing.	Entretanto, a pesquisa de novos tratamentos está em curso.							
53	The earlier PH is treated, the easier it is to control.	O PH mais cedo for tratado, o que é mais fácil de controlar.	3				2		
54	Treatments include medicines, procedures, and other therapies.	Os tratamentos incluem medicamentos, procedimentos e outras terapias.							
55	These treatments can relieve PH symptoms and slow the progress of the disease.	Estes tratamentos podem aliviar os sintomas de PH e para retardar o progresso da doença.				2	2		
56	Lifestyle changes also can help control symptoms.	Lifestyle mudanças também podem ajudar a controlar os sintomas.					3		
57	Lifestyle changes	lifestyle mudanças					3	1	

58	- Monitor your weight: weigh yourself at the same time each day and record your weight in a diary or on a calendar.	- Monitore seu peso: pese-se ao mesmo tempo a cada dia e gravar o seu peso em um diário ou em um calendário.	1				2		
59	- Lose weight if you are overweight (follow your doctor's recommendations).	- Perca peso se você está acima do peso (siga as recomendações do seu médico).		1					
60	- Have annual health check-ups with your primary care physician.	- Tenha anuais de saúde check-ups com o seu médico de cuidados primários.	2						
61	- Activity guidelines: restrict lifting, pushing, or shoving to less than 20 pounds, since these activities increase the pressure in your arteries and lungs; walking is important to keep your muscles strong and improve circulation.	- Diretrizes Atividade: Restringir o levantamento, empurrando ou empurrando para menos de 20 quilos, uma vez que estas atividades aumentam a pressão em suas artérias e pulmões; caminhar é importante manter seus músculos fortes e melhorar a circulação.	2			1	1	1	
62	However, more strenuous forms of aerobic exercise should not be done regularly since your lungs may not be able to keep up with the oxygen demands placed on your body during these activities.	No entanto, as formas mais vigorosas de exercício aeróbico não deve ser feito regularmente desde os seus pulmões podem não ser capaz de manter-se com as demandas de oxigênio colocados em seu corpo durante essas atividades.	1				3		
63	Any activities that cause shortness of breath, dizziness, or chest pain should be stopped immediately.	Todas as atividades que causam falta de ar, tonturas ou dor no peito deve ser interrompida imediatamente.	1						
64	Contact your doctor if any of these symptoms occur.	Contate o seu médico se algum destes sintomas ocorrer.							
65	- Do not smoke or chew tobacco.	- Não fumar ou mastigar tabaco.							
66	- Avoid or reduce alcohol intake.	- Evitar ou reduzir a ingestão de álcool.							
67	- Avoid pregnancy; use contraceptives to prevent pregnancy.	- Evitar a gravidez; usam contraceptivos para prevenir a gravidez.					2		
68	- Maintain close contact with your health care providers.	- Manter contato próximo com os seus prestadores de cuidados de saúde.							
69	When to call your doctor?	Quando chamar o médico?							
70	- Weight gain — 2 pounds in one day or 5 pounds in one week	- Ganho de peso - 2 quilos em um dia ou 5 quilos em uma semana							
71	- Swelling in your ankles, legs, or abdomen that has become worse	- Inchaço nos tornozelos, pernas ou abdômen, que se tornou pior							
72	- Shortness of breath that has become worse, especially if you wake up short of breath	- Falta de ar que se tornou pior, especialmente se você acordar com falta de ar							
73	- Extreme fatigue or decreased tolerance for activity	- Fadiga extrema ou diminuição da tolerância à atividade							

74	- A respiratory infection or a cough that has become worse	- A infecção respiratória ou uma tosse que se tornou pior					1		
75	- Fast heart rate (around 120 beats per minute)	- Frequência cardíaca rápida (cerca de 120 batimentos por minuto)							
76	- Episodes of chest pain or discomfort with exertion that are not relieved with rest	- Os episódios de dor ou desconforto no peito com o esforço que não são aliviados com repouso				1			
77	- Difficulty breathing with regular activities or at rest	- Dificuldade em respirar com atividades regulares ou em repouso							
78	- Restlessness, confusion	- Inquietação, confusão							
79	- Constant dizziness or lightheadedness	- Tonturas ou vertigens Constante	1					1	
80	- Nausea or poor appetite	- Náuseas ou falta de apetite							
81	When to go to the emergency department?	Quando ir ao departamento de emergência?		1					
82	- Hickman catheter complications with intravenous Flolan, such as infection, catheter displacement/leakage of solution or blood, IV pump malfunction	- Hickman complicações do cateter intravenoso com Flolan, como infecção, deslocamento do cateter / vazamento de solução ou de sangue, IV bomba avaria	3						
83	- New chest pain or discomfort that is severe, unexpected, and is accompanied by shortness of breath, sweating, nausea, or weakness	- Dor no peito ou desconforto Nova que é grave, inesperado, e é acompanhada de falta de ar, sudorese, náuseas, ou fraqueza	2					1	
84	- Fast, sustained heart rate (more than 120 to 150 beats per minute) — especially if you are short of breath	- Rápido, sustentada frequência cardíaca (mais de 120 a 150 batimentos por minuto) - especialmente se você estiver com falta de ar	2						
85	- Shortness of breath that is NOT relieved by rest	- Falta de ar que não seja aliviada pelo repouso						1	
86	- Sudden weakness or paralysis in your arms or legs	- Fraqueza súbita ou paralisia em seus braços ou pernas							
87	- Sudden onset of severe headache	- O início súbito de dor de cabeça severa				1			
88	- Fainting spells with loss of consciousness	- Desmaios com perda de consciência							
89	Preeclampsia and Eclampsia	A pré-eclâmpsia e eclâmpsia				1			
90	Preeclampsia (toxemia) and eclampsia are conditions that can develop during week 20 or later of your pregnancy.	A pré-eclâmpsia (toxemia) e eclâmpsia são condições que podem se desenvolver durante a semana 20 ou mais tarde, de sua gravidez.			1				
91	With preeclampsia, you have high blood pressure and protein in your urine.	Com a pré-eclâmpsia, você tem pressão arterial elevada e proteína na urina.							
92	Preeclampsia progresses to eclampsia if you have a seizure.	A pré-eclâmpsia progride para eclâmpsia, se você tem uma convulsão.			1				
93	This does not include women with a known seizure disorder.	Isso não inclui as mulheres com um distúrbio convulsivo conhecido.		1					

94	These conditions can create mild to life-threatening health problems for you and your unborn baby.	Estas condições podem criar leve a problemas de saúde com risco de vida para você e seu bebê por nascer.	3						
95	You will need tests 1 to 2 times a week to check your condition.	Você vai precisar de testes de 1 a 2 vezes por semana para verificar a sua condição.							
96	Tests include blood pressure checks, urine and blood tests, and fetal monitoring.	Os testes incluem verificações de pressão arterial, urina e exames de sangue e monitorização fetal.							
97	Rest:	descanso:						1	
98	- Your obstetrician may tell you to rest more often if you have mild symptoms of preeclampsia.	- Seu obstetra pode dizer-lhe para descansar mais frequentemente se você tem sintomas leves de pré-eclâmpsia.							
99	You may need total bedrest if you have more severe symptoms.	Você pode precisar de repouso absoluto total, se você tem sintomas mais graves.			1				
100	Try to lie on your left side.	Tente estar no seu lado esquerdo.						3	

vi. “Textos Mais Frequentes” – Original e tradução de referência

	Original_ENUS	Tradução_PT
1	- Accidents and safety:.	- Acidentes e segurança:.
2	- Adequate/inadequate feeding.	- Alimentação adequada/inadequada.
3	- Bottle weaning:.	- Desmame do biberão:.
4	- Child's development and temperament:	- Desenvolvimento e temperamento da criança:
5	- Child's development, temperament and relationship:	- Desenvolvimento, temperamento e relacionamento da criança:
6	- Child's reaction to strangers:	- Reação da criança a estranhos:
7	- Cognitive development, laterality:.	- Desenvolvimento cognitivo, lateralidade:.
8	- Cognitive development:.	- Desenvolvimento cognitivo:.
9	- Complete physical examination was performed and the child's state is normal.	- Foi efetuado o exame físico completo e tudo se encontra normal.
10	- Dentition and oral hygiene:.	- Dentição e higiene oral:.
11	- Dentition and teeth brushing:	- Dentição e lavagem dos dentes:
12	- Dentition and teeth brushing:.	- Dentição e lavagem dos dentes:.
13	- Dentition, teeth brushing and use of dental floss:.	- Dentição, lavagem dos dentes e utilização de fio dentário:.
14	- Dentition:	- Dentição:
15	- Dentition:.	- Dentição:.
16	- Discovery of sexual orientation:.	- Descoberta da identidade sexual:.
17	- Eating habits:.	- Hábitos alimentares
18	- Gait:.	- Marcha:.
19	- General aspects:.	- Aspetos gerais:.
20	- Harmful consumption:.	- Consumos nocivos:.
21	- Hearing/Language:	- Audição / Linguagem:
22	- Hearing:	- Audição:
23	- Issues related to accidents and security were/were not discussed.	- Foram/não foram abordadas questões acerca de acidentes e segurança.
24	- Issues related to the child's nutrition were/were not discussed:.	- Foram/não foram discutidos aspetos acerca da alimentação da criança.
25	- Issues related to the introduction of new food into a child's diet were/were not discussed.	- Foi/não foi discutida a introdução de novos alimentos na alimentação da criança.
26	- Language development:	- Desenvolvimento da linguagem:
27	- Language:	- Língua:
28	- Language:.	- Língua:.
29	- Learning of daily routines and rules:	- Aprendizagem de regras e rotinas da vida diária:
30	- Lipid profile:.	- Perfil lipídico:.
31	- Main diet errors:	- Principais erros alimentares:
32	- New skills and acquisitions (motor and language):	- Novas capacidades e aquisições (motoras e de linguagem):
33	- Normal development for age.	- Desenvolvimento normal para a idade.
34	- Normal/abnormal hearing.	- Audição sem/com alterações.
35	- Normal/abnormal vision.	- Visão sem/com alterações.
36	- Oral health:.	- Saúde oral:.
37	- Orthopedic problems:	- Problemas ortopédicos:
38	- Pacifier weaning:.	- Desmame da chupeta:.
39	- Participates/does not participate in sports and cultural activities.	- Participa/não participa em desportos e atividades culturais.
40	- Personality assertion, stubbornness, negativism.	- Afirmação da personalidade, birras, negativismo.
41	- Physical exercise, outdoor activities, leisure and intellectual stimulation:.	- Exercício físico, atividades ao ar livre, atividades de lazer e estímulo intelectual:.
42	- Playing, drawing, television habits, going to bed ritual:	- Brincar, desenhar, hábitos de televisão, ritual de ir para a cama:

43	- Preparation for admission to school, prevention of school failure, attitude towards school and studying:.	- Preparação da entrada para a escola, prevenção do insucesso escolar, atitude em relação à escola e ao estudo:.
44	- Psychomotor development.	- Desenvolvimento psicomotor.
45	- Psychomotor development:.	- Desenvolvimento psicomotor:.
46	- Pubertal stage:.	- Estádio pubertário:.
47	- Puberty:.	- Puberdade:.
48	- Reading habits:.	- Hábitos de leitura:.
49	- Relationship with other children:.	- Relacionamento com outras crianças:.
50	- Sleeping/resting patterns, nocturnal enuresis:.	- Hábitos de sono/repouso, enurese noturna:.
51	- Sphincter control:	- Controlo de esfíncteres:
52	- Temperament - negativism, stubbornness, jealousy, rivalry:	- Temperamento - negativismo, birras, ciúmes, rivalidade:
53	- Temperament and relationship - independence, separation anxiety, night terrors:	- Temperamento e relacionamento - independência, ansiedade da separação, terrores noturnos:
54	- Toys:	- Brinquedos:
55	- Vaccination:.	- Vacinação:.
56	- Vision:.	- Visão:.
57	Fr nasogastric tube, for continuous drainage to collection bag with clamp / feeding by...	Sonda nasogástrica ____ Fr de drenagem contínua para saco coletor com clamp/alimentação por...
58	Fr urinary catheter, for continuous drainage to collection bag with clamp.	Sonda vesical de ____ Fr de drenagem contínua para saco coletor com clamp.
59	Cardiotocography (CTG)	Cardiotocografia (CTG)
60	Discharge conditions:	Condições da alta:
61	Admission date:	Data de admissão:
62	Discharge date:	Data de alta:
63	Admission diagnosis:	Diagnóstico de admissão:
64	Discharge diagnosis:	Diagnóstico de alta:
65	... weeks pregnant.	Grávida com ... semanas de gestação.
66	Reason for visit:	Motivo da consulta:
67	Para	Para
68	- Leisure occupations:.	- Ocupações de tempos livres:.
69	- Sleeping/resting patterns:.	- Hábitos de sono/repouso:.
70	- Social and cultural interests:.	- Interesses culturais e sociais:.
71	- Statural and ponderal growth:.	- Desenvolvimento estatural-ponderal:.
72	weeks of high-risk pregnancy.	Gravidez de alto risco, com a duração de.....semanas.
73	weeks of pregnancy without complications.	Gravidez sem intercorrências, com a duração de.....semanas.
74	? year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Paciente de ... anos que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com repouso.
75	? year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	Paciente com ... anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.
76	... year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Paciente de ... anos que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com repouso.
77	... year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	Paciente com ... anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.
78	weeks of high-risk pregnancy.	Gravidez de alto risco, com a duração de ...

		semanas.
79	weeks of pregnancy without complications.	Gravidez sem intercorrências, com a duração de ... semanas.
80	48 hours.	48 horas.
81	Abdominal pain since...	Dor abdominal desde...
82	Abnormal genitourinary exam - Female.	Exame geniturinário com alterações – mulher
83	Abnormal genitourinary exam - Male.	Exame geniturinário com alterações – homem
84	Abnormal haematologic system.	Sistema hematológico com alterações.
85	Absence / Temporary absence from clinical service due to.....	Ausência/ausência temporária deste serviço clínico devido a...
86	Admitted for: labs/antibiotics/therapies	Internado para: análises/antibióticos/tratamentos
87	Foetal heart rate (FHR):	Frequência cardíaca fetal (FCF):
88	Gravida; ...	Gesta; ...
89	Number of appointments.....	Número de consultas...
90	Number of appointments.....	Número de consultas...
91	Procedures executed in clinic:	Procedimentos efetuados durante a estadia na clínica:
92	Procedures executed in hospital:	Procedimentos efetuados durante a estadia no hospital:
93	Procedures executed in primary care center:	Procedimentos efetuados durante a estadia no centro de saúde:
94	year old patient, smoker, with a history of arterial hypertension, has retrosternal oppressive pain with radiation to the left arm and mandible, associated with sudoresis and nausea, onset date/time , not relieved by rest.	Doente com __ anos de idade, fumador, com antecedentes de hipertensão arterial, com dor opressiva retroesternal com irradiação ao braço esquerdo e mandíbula, associada a sudorese e náuseas, com início _____ , sem alívio pelo repouso.

vii. “Textos Mais Frequentes” – Avaliação da tradução do Moses for Mere Mortals (ALL)

	ORIGINAL_EN	MMM_PT (ALL)	Gra	Ter	Po	Om	Tra	Mai	Coe
1	- Accidents and safety:.	- acidentes e segurança:.						1	
2	- Adequate/inadequate feeding.	- alimentação adequada/não adequada.	1					1	
3	- Bottle weaning:.	- Desmame do biberão:.							
4	- Child's development and temperament:	- desenvolvimento e Temperamento da criança:						1	
5	- Child's development, temperament and relationship:	- desenvolvimento, Temperamento e relacionamento da criança:						1	
6	- Child's reaction to strangers:	- reacção da criança a estranhos:						1	
7	- Cognitive development, laterality:.	- desenvolvimento cognitivo, Lateralidade:.						1	
8	- Cognitive development:.	- desenvolvimento cognitivo:.						1	
9	- Complete physical examination was performed and the child's state is normal.	- foi efetuado o exame físico completo e tudo se encontra normal.						1	
10	- Dentition and oral hygiene:.	- Dentição e higiene oral:.	1						
11	- Dentition and teeth brushing:	- Dentição e lavagem dos dentes:							
12	- Dentition and teeth brushing:.	- Dentição e lavagem dos dentes:.							
13	- Dentition, teeth brushing and use of dental floss:.	- Dentição, lavagem dos dentes e utilização de fio dentário:.							

14	- Dentition:	- Dentição:							
15	- Dentition:.	- Dentição:.							
16	- Discovery of sexual orientation:.	- descoberta da identidade sexual:.						1	
17	- Eating habits:.	- hábitos alimentares:.						1	
18	- Gait:.	- postura:.		3				1	
19	- General aspects:.	- aspectos gerais:.						1	
20	- Harmful consumption:.	- consumos nocivos:.						1	
21	- Hearing/Language:	- audição/linguagem:						1	
22	- Hearing:	- audição:						1	
23	- Issues related to accidents and security were/were not discussed.	- questões relacionadas com acidentes e segurança foram/não foram discutidos.						1	
24	- Issues related to the child's nutrition were/were not discussed:.	- aspectos acerca da alimentação da criança foram/não foram:.						1	
25	- Issues related to the introduction of new food into a child's diet were/were not discussed.	- introduzidos novos alimentos na alimentação da criança foram/não foram discutidos.				3	1	1	
26	- Language development:	- desenvolvimento da linguagem:						1	
27	- Language:	- língua:						1	

28	- Language:.	- língua:.						1	
29	- Learning of daily routines and rules:	- aprendizagem de regras e rotinas na vida diária:						1	
30	- Lipid profile:.	- perfil lipídico:.						1	
31	- Main diet errors:	- principais erros alimentares:						1	
32	- New skills and acquisitions (motor and language):	- novas capacidades e aquisições (motoras e de linguagem):						1	
33	- Normal development for age.	- desenvolvimento normal para a idade.						1	
34	- Normal/abnormal hearing.	- audição Sem/com alterações.						1	
35	- Normal/abnormal vision.	- visão Sem/com alterações.						1	
36	- Oral health:.	- saúde oral:.						1	
37	- Orthopedic problems:	- problemas ortopédicos:						1	
38	- Pacifier weaning:.	- Desmame da chupeta:.							
39	- Participates/does not participate in sports and cultural activities.	- pratica/não pratica desporto e actividades culturais.					1	1	
40	- Personality assertion, stubbornness, negativism.	- Afirmação da personalidade, birras, negativismo.							
41	- Physical exercise, outdoor activities, leisure and intellectual stimulation:.	- exercício físico, ar livre, actividades de lazer e estímulo intelectual:.	2					1	

42	- Playing, drawing, television habits, going to bed ritual:	- brincar, desenhar, hábitos de televisão, ritual de ir para a cama:						1	
43	- Preparation for admission to school, prevention of school failure, attitude towards school and studying:.	- preparação da entrada para a escola, prevenção do Insucesso escolar, postura na escola e no estudo:.						1	
44	- Psychomotor development.	- desenvolvimento psico-motor:	1		1			1	
45	- Psychomotor development:.	- desenvolvimento psico-motor:.	1					1	
46	- Pubertal stage:.	- Estádio pubertário:.							
47	- Puberty:.	- puberdade:.						1	
48	- Reading habits:.	- hábitos de leitura:.						1	
49	- Relationship with other children:.	- relacionamento com as outras crianças:.						1	
50	- Sleeping/resting patterns, nocturnal enuresis:.	- hábitos de sono/repouso, enurese nocturna:.						1	
51	- Sphincter control:	- controlo de Esfíncteres:						1	
52	- Temperament - negativism, stubbornness, jealousy, rivalry:	- Temperamento - negativismo, birras, ciúmes, rivalidade:							
53	- Temperament and relationship - independence, separation anxiety, night terrors:	- Temperamento e relacionamento - independência, ansiedade da separação, terrores nocturnos:							
54	- Toys:	- brinquedos:						1	
55	- Vaccination:.	- Vacinação:.							

56	- Vision:.	- visão:.						1	
57	Fr nasogastric tube, for continuous drainage to collection bag with clamp / feeding by...	Fr sonda nasogástrica, em drenagem livre para saco colector com clamp/alimentação por...	1						
58	Fr urinary catheter, for continuous drainage to collection bag with clamp.	Fr catéter urinário, em drenagem livre para saco colector/clampada para alimentação por ú	1			2			1
59	Cardiotocography (CTG)	Cardiotocografia (CTG)							
60	Discharge conditions:	Condições da alta:							
61	Admission date:	Data de admissão:							
62	Discharge date:	Data de alta:							
63	Admission diagnosis:	Diagnóstico de admissão:							
64	Discharge diagnosis:	Diagnóstico final:							
65	... weeks pregnant.	Grávida de semanas com... gestação.	2			2			
66	Reason for visit:	Motivo da consulta:							
67	Para	For		3					

68	- Leisure occupations:.	- Ocupações de tempos livres:.							
69	- Sleeping/resting patterns:.	- hábitos de sono/repouso:.						1	
70	- Social and cultural interests:.	- interesses culturais e sociais:.							
71	- Statural and ponderal growth:.	- desenvolvimento estatura-ponderal:.							
72	weeks of high-risk pregnancy.	 semanas de gravidez de alto risco.							
73	weeks of pregnancy without complications.	Gravidez sem intercorrências, com a duração de semanas.							
74	? year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Paciente? anos que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com descanso.							
75	? year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	? anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.						3	
76	... year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Paciente de... anos que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com descanso.							
77	... year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	Paciente com... anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.						3	
78	weeks of high-risk pregnancy.	 semanas de gravidez de alto risco.							
79	weeks of pregnancy without complications.	 semanas de gravidez sem intercorrências.							
80	48 hours.	48 horas.							

81	Abdominal pain since...	Dor abdominal desde...							
82	Abnormal genitourinary exam - Female.	Exame geniturinário com alterações – mulher.							
83	Abnormal genitourinary exam - Male.	Exame geniturinário com alterações – homem.							
84	Abnormal haematologic system.	Anormal haematologic sistema.	1				2		1
85	Absence / Temporary absence from clinical service due to.....	Ausência/ausência temporária deste serviço clínico devido a							
86	Admitted for: labs/antibiotics/therapies	Admitido para: análises/antibióticos/terapias							
87	Foetal heart rate (FHR):	Frequência cardíaca fetal (FTC):					2		
88	Gravida; ...	Gesta; ...							
89	Number of appointments.....	Número de consultas							
90	Number of appointments.....	Número de consultas							
91	Procedures executed in clinic:	Procedimentos efetuados durante a estadia na clínica:							
92	Procedures executed in hospital:	Procedimentos efetuados durante a estadia no hospital:							
93	Procedures executed in primary care center:	Procedimentos efetuados durante a estadia no Centro de saúde:						1	
94	year old patient, smoker, with a history of arterial hypertension, has retrosternal oppressive pain with radiation to the left arm and mandible, associated with sudoresis and nausea, onset date/time , not	Anos de idade, fumador, com antecedentes de hipertensão arterial, com dor opressiva retroesternal com irradiação ao braço esquerdo e mandíbula, associada a sudorese e náuseas, com início, sem alívio pelo repouso.						1	

	relieved by rest.									
--	-------------------	--	--	--	--	--	--	--	--	--

viii. “Textos Mais Frequentes” – Avaliação da tradução do Moses for Mere Mortals (EMEA)

	ORIGINAL_EN	MMM_PT (EMEA)	Gr a	Ter	Po	Om	Tra	Mai	Coe
1	- Accidents and safety:.	- acidentes e segurança:.						1	
2	- Adequate/inadequate feeding.	- alimentação adequada/não adequada.	1					1	
3	- Bottle weaning:.	- desmame do biberão:.						1	
4	- Child's development and temperament:	- desenvolvimento e Temperamento da criança:						1	
5	- Child's development, temperament and relationship:	- desenvolvimento, Temperamento e relacionamento da criança:						1	
6	- Child's reaction to strangers:	- reacção da criança a estranhos:						1	
7	- Cognitive development, laterality:.	- desenvolvimento cognitivo, Lateralidade:.						1	
8	- Cognitive development:.	- desenvolvimento cognitivo:.						1	
9	- Complete physical examination was performed and the child's state is normal.	- foi efetuado o exame físico completo e tudo se encontra normal. :			1			1	
10	- Dentition and oral hygiene:.	- dentição e higiene oral:.	1					1	
11	- Dentition and teeth brushing:	- dentição e lavagem dos dentes:						1	
12	- Dentition and teeth brushing:.	- dentição e lavagem dos dentes:.						1	
13	- Dentition, teeth brushing and use of dental floss:.	- dentição, lavagem dos dentes e utilização de fio dentário:.						1	

14	- Dentition:	- dentição:							1	
15	- Dentition:.	- dentição:.							1	
16	- Discovery of sexual orientation:.	- descoberta da identidade sexual:.							1	
17	- Eating habits:.	- hábitos alimentares:.							1	
18	- Gait:.	- postura: .		3					1	
19	- General aspects:.	- aspectos gerais:.							1	
20	- Harmful consumption:.	- consumos nocivos:.							1	
21	- Hearing/Language:	- audição /linguagem:							1	
22	- Hearing:	- audição:							1	
23	- Issues related to accidents and security were/were not discussed.	- questões relacionadas com acidentes e segurança foram/não foram discutidos .	1						1	
24	- Issues related to the child's nutrition were/were not discussed:.	- aspectos acerca da alimentação da criança foram/não foram discutidos:.							1	
25	- Issues related to the introduction of new food into a child's diet were/were not discussed.	- introduzidos novos alimentos na alimentação da criança foram/não foram discutidos.				3	1		1	
26	- Language development:	- desenvolvimento da linguagem:							1	
27	- Language:	- língua:							1	

28	- Language:.	- língua:.						1	
29	- Learning of daily routines and rules:	- aprendizagem de regras e rotinas na vida diária:						1	
30	- Lipid profile:.	- perfil lipídico:.						1	
31	- Main diet errors:	- principais erros alimentares:						1	
32	- New skills and acquisitions (motor and language):	- novas capacidades e aquisições (motoras e de linguagem):						1	
33	- Normal development for age.	- desenvolvimento normal para a idade.						1	
34	- Normal/abnormal hearing.	- audição sem/com alterações. 1			1			1	
35	- Normal/abnormal vision.	- visão sem/com alterações. 1			1			1	
36	- Oral health:.	- saúde oral:.						1	
37	- Orthopedic problems:	- problemas ortopédicos:						1	
38	- Pacifier weaning:.	- desmame da chupeta:.						1	
39	- Participates/does not participate in sports and cultural activities.	- pratica/não pratica desporto e actividades culturais.					1	1	
40	- Personality assertion, stubbornness, negativism.	- afirmação da personalidade, birras, negativismo.						1	
41	- Physical exercise, outdoor activities, leisure and intellectual stimulation:.	- exercício físico, ar livre, actividades de lazer e estímulo intelectual:.	2					1	

42	- Playing, drawing, television habits, going to bed ritual:	- brincar, desenhar, hábitos de televisão, ritual de ir para a cama:						1	
43	- Preparation for admission to school, prevention of school failure, attitude towards school and studying:.	- preparação da entrada para a escola, prevenção do insucesso escolar, postura na escola e no estudo:.						1	
44	- Psychomotor development.	- desenvolvimento psico-motor:	1					1	
45	- Psychomotor development:.	- desenvolvimento psico-motor:.	1					1	
46	- Pubertal stage:.	- estágio pubertário:.						1	
47	- Puberty:.	- puberdade:.						1	
48	- Reading habits:.	- hábitos de leitura:.						1	
49	- Relationship with other children:.	- relacionamento com as outras crianças:.						1	
50	- Sleeping/resting patterns, nocturnal enuresis:.	- hábitos de sono/repouso, enurese nocturna:.						1	
51	- Sphincter control:	- controlo de Esfíncteres:						1	
52	- Temperament - negativism, stubbornness, jealousy, rivalry:	- Temperamento - negativismo, birras, ciúmes, rivalidade:							
53	- Temperament and relationship - independence, separation anxiety, night terrors:	- Temperamento e relacionamento - independência, ansiedade da separação, terrores nocturnos:							
54	- Toys:	- brinquedos:						1	
55	- Vaccination:.	- vacinação:.						1	

56	- Vision:.	- visão:.						1	
57	Fr nasogastric tube, for continuous drainage to collection bag with clamp / feeding by...	FR sonda nasogástrica, em drenagem livre para saco colector com clamp/alimentação por...	1						
58	Fr urinary catheter, for continuous drainage to collection bag with clamp.	FR catéter urinário, em drenagem livre para saco colector/clampada para alimentação por g	1			2			1
59	Cardiotocography (CTG)	Cardiotocografia (CTG)							
60	Discharge conditions:	Condições da alta:							
61	Admission date:	Data de admissão:							
62	Discharge date:	Data de alta:							
63	Admission diagnosis:	Diagnóstico de admissão:							
64	Discharge diagnosis:	Diagnóstico final:							
65	... weeks pregnant.	Semanas com... gestação.					2		
66	Reason for visit:	Motivo da consulta:							
67	Para	For		3					

68	- Leisure occupations:.	- Ocupações de tempos livres:.							
69	- Sleeping/resting patterns:.	- hábitos de sono/repouso:.						1	
70	- Social and cultural interests:.	- interesses culturais e sociais:.						1	
71	- Statural and ponderal growth:.	- desenvolvimento estatura-ponderal:.						1	
72	weeks of high-risk pregnancy.	 semanas de gravidez de alto risco.							
73	weeks of pregnancy without complications.	Gravidez sem intercorrências, com a duração de semanas.							
74	? year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	?, que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com descanso.				2			
75	? year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	? anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.					3		
76	... year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Paciente de... anos que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com descanso.							
77	... year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	Paciente com... anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.					3		
78	weeks of high-risk pregnancy.	Gravidez de alto risco, com a duração de semanas.							
79	weeks of pregnancy without complications.	 semanas de gravidez sem intercorrências.							
80	48 hours.	0					3		

81	Abdominal pain since...	Dor abdominal desde...							
82	Abnormal genitourinary exam - Female.	Exame geniturinário com alterações – mulher.							
83	Abnormal genitourinary exam - Male.	Exame geniturinário com alterações – homem.							
84	Abnormal haematologic system.	Com alterações sistema hematológico.	2						
85	Absence / Temporary absence from clinical service due to.....	Ausência/ausência temporária deste serviço clínico devido a							
86	Admitted for: labs/antibiotics/therapies	Admitido para: análises/antibióticos/terapêuticas							
87	Foetal heart rate (FHR):	Frequência cardíaca fetal (FIC):					2		
88	Gravida; ...	Gesta; ...							
89	Number of appointments.....	Número de consultas							
90	Number of appointments.....	Número de consultas							
91	Procedures executed in clinic:	Procedimentos efetuados durante a estadia na clínica:							
92	Procedures executed in hospital:	Procedimentos efetuados durante a estadia no hospital:							
93	Procedures executed in primary care center:	Procedimentos efetuados durante a estadia no Centro de saúde:						1	
94	year old patient, smoker, with a history of arterial hypertension, has retrosternal oppressive pain with radiation to the left arm and mandible, associated with sudoresis and nausea, onset date/time , not	Anos de idade, fumador, com antecedentes de hipertensão arterial, com dor opressiva retroesternal com irradiação ao braço esquerdo e mandíbula, associada a sudorese e náuseas, com início, sem alívio pelo repouso.						1	

	relieved by rest.									
--	-------------------	--	--	--	--	--	--	--	--	--

ix. “Textos Mais Frequentes” – Avaliação da tradução do Moses for Mere Mortals (Europarl)

	ORIGINAL_EN	MMM_PT (Europarl)	Gr a	Ter	Po	Om	Tra	Mai	Coe
1	- Accidents and safety:.	- acidentes e segurança:.						1	
2	- Adequate/inadequate feeding.	- alimentação adequada/não adequada.	1					1	
3	- Bottle weaning:.	- desmame do biberão:.						1	
4	- Child's development and temperament:	- desenvolvimento e temperamento da criança:						1	
5	- Child's development, temperament and relationship:	- desenvolvimento, temperamento e relacionamento da criança:						1	
6	- Child's reaction to strangers:	- reacção da criança a estranhos:						1	
7	- Cognitive development, laterality:.	- desenvolvimento cognitivo, Lateralidade:.						1	
8	- Cognitive development:.	- desenvolvimento cognitivo:.						1	
9	- Complete physical examination was performed and the child's state is normal.	- foi efectuado exame físico completo e tudo se encontra normal.						1	
10	- Dentition and oral hygiene:.	- dentição e higiene oral:.	1					1	
11	- Dentition and teeth brushing:	- dentição e lavagem dos dentes:						1	
12	- Dentition and teeth brushing:.	- dentição e lavagem dos dentes:.						1	
13	- Dentition, teeth brushing and use of dental floss:.	- dentição, lavagem dos dentes e utilização de fio dentário:.						1	

14	- Dentition:	- dentição:							1	
15	- Dentition:.	- dentição:.							1	
16	- Discovery of sexual orientation:.	- descoberta da identidade sexual:.							1	
17	- Eating habits:.	- hábitos alimentares:.							1	
18	- Gait:.	- postura:.		3					1	
19	- General aspects:.	- aspectos gerais:.							1	
20	- Harmful consumption:.	- consumos nocivos:.							1	
21	- Hearing/Language:	- audição/linguagem:							1	
22	- Hearing:	- audição:							1	
23	- Issues related to accidents and security were/were not discussed.	- questões acerca de acidentes e segurança foram/não foram discutidos.	1						1	
24	- Issues related to the child's nutrition were/were not discussed:.	- aspectos acerca da alimentação da criança foram/não foram discutidos:.							1	
25	- Issues related to the introduction of new food into a child's diet were/were not discussed.	- introduzidos novos alimentos na alimentação da criança foram/não foram discutidos.				3	1	1		
26	- Language development:	- desenvolvimento da linguagem:							1	
27	- Language:	- língua:							1	

28	- Language:.	- língua:.							1	
29	- Learning of daily routines and rules:	- aprendizagem de regras e rotinas na vida diária:							1	
30	- Lipid profile:.	- perfil lipídico:.							1	
31	- Main diet errors:	- principais erros alimentares:							1	
32	- New skills and acquisitions (motor and language):	- novas capacidades e aquisições (motoras e de linguagem):							1	
33	- Normal development for age.	- desenvolvimento normal para a idade.							1	
34	- Normal/abnormal hearing.	- audição sem/com alterações.							1	
35	- Normal/abnormal vision.	- visão sem/com alterações.							1	
36	- Oral health:.	- saúde oral:.							1	
37	- Orthopedic problems:	- problemas ortopédicos:							1	
38	- Pacifier weaning:.	- desmame da chupeta:.							1	
39	- Participates/does not participate in sports and cultural activities.	- pratica/não pratica desporto e actividades culturais.						1	1	
40	- Personality assertion, stubbornness, negativism.	- afirmação da personalidade, birras, negativismo.							1	
41	- Physical exercise, outdoor activities, leisure and intellectual stimulation:.	- exercício físico, ar livre, actividades de lazer e estímulo intelectual:.	2						1	

42	- Playing, drawing, television habits, going to bed ritual:	- brincar, desenhar, hábitos de televisão, ritual de ir para a cama:						1	
43	- Preparation for admission to school, prevention of school failure, attitude towards school and studying:.	- preparação da entrada para a escola, prevenção do insucesso escolar, postura na escola e no estudo:.						1	
44	- Psychomotor development.	- desenvolvimento psico-motor:	1		1			1	
45	- Psychomotor development:.	- desenvolvimento psico motor:.	1					1	
46	- Pubertal stage:.	- estágio pubertário:.						1	
47	- Puberty:.	- puberdade:.						1	
48	- Reading habits:.	- hábitos de leitura:.						1	
49	- Relationship with other children:.	- relacionamento com as outras crianças:.						1	
50	- Sleeping/resting patterns, nocturnal enuresis:.	- hábitos de sono/repouso, enurese noturna:.						1	
51	- Sphincter control:	- controlo de Esfíncteres:						1	
52	- Temperament - negativism, stubbornness, jealousy, rivalry:	- temperamento - negativismo, birras, ciúmes, rivalidade:						1	
53	- Temperament and relationship - independence, separation anxiety, night terrors:	- temperamento e relacionamento - independência, ansiedade da separação, terrores noturnos:						1	
54	- Toys:	- brinquedos:						1	
55	- Vaccination:.	- vacinação:.						1	

56	- Vision:.	- visão:.						1	
57	Fr nasogastric tube, for continuous drainage to collection bag with clamp / feeding by...	FR sonda nasogástrica, em drenagem livre para saco colector com clamp/alimentação por...	1						
58	Fr urinary catheter, for continuous drainage to collection bag with clamp.	FR catéter urinário, em drenagem livre para saco colector/clampada/para alimentação por d	1			2			1
59	Cardiotocography (CTG)	Cardiotocografia (CTG)							
60	Discharge conditions:	Condições da alta:							
61	Admission date:	Data de admissão:							
62	Discharge date:	Data de alta:							
63	Admission diagnosis:	Diagnóstico de admissão:							
64	Discharge diagnosis:	Diagnóstico de alta:							
65	... weeks pregnant.	Semanas com ... gestação.					2		
66	Reason for visit:	Motivo da consulta:							
67	Para	Por					3		

68	- Leisure occupations:.	- ocupações de tempos livres:.						1	
69	- Sleeping/resting patterns:.	- hábitos de sono/repouso:.						1	
70	- Social and cultural interests:.	- interesses culturais e sociais:.						1	
71	- Statural and ponderal growth:.	- desenvolvimento estatura-ponderal:.						1	
72	weeks of high-risk pregnancy.	 semanas de gravidez de alto risco.							
73	weeks of pregnancy without complications.	Gravidez sem intercorrências, com a duração de semanas.							
74	? year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Ano? o paciente que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com descanso.	2						
75	? year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	? anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.					3		
76	... year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Paciente de... anos que apresenta desconforto retroesternal induzido por esforço (fazer a cama, subir escadas, lavar pratos) e aliviado com descanso.							
77	... year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	Paciente com... anos de idade, fumador, com antecedentes familiares de hipertensão, com desconforto retroesternal a irradiar para o braço esquerdo e maxilar associado a falta de ar e náuseas, aliviado com repouso.					3		
78	weeks of high-risk pregnancy.	 semanas de gravidez de alto risco.							
79	weeks of pregnancy without complications.	 semanas de gravidez sem intercorrências.							
80	48 hours.	48 horas.							

81	Abdominal pain since...	Dor abdominal desde...							
82	Abnormal genitourinary exam - Female.	Exame geniturinário com alterações – mulher.							
83	Abnormal genitourinary exam - Male.	Exame geniturinário com alterações – Homem.							1
84	Abnormal haematologic system.	Com alterações sistema hematológico:	1						
85	Absence / Temporary absence from clinical service due to.....	Ausência/ausência temporária deste serviço clínico devido a							
86	Admitted for: labs/antibiotics/therapies	Admitidos por: labs/antibióticos/tratamentos					2		
87	Foetal heart rate (FHR):	Frequência cardíaca fetal (FTC):					2		
88	Gravida; ...	Gesta; ...							
89	Number of appointments.....	Número de consultas							
90	Number of appointments.....	Número de consultas							
91	Procedures executed in clinic:	Procedimentos efetuados durante a estadia na clínica:							
92	Procedures executed in hospital:	Procedimentos efetuados durante a estadia no hospital:							
93	Procedures executed in primary care center:	Procedimentos efetuados durante a estadia no centro de saúde:							
94	year old patient, smoker, with a history of arterial hypertension, has retrosternal oppressive pain with radiation to the left arm and mandible, associated with sudoresis and nausea, onset date/time , not	Anos de idade, fumador, com antecedentes de hipertensão arterial, com dor opressiva retroesternal com irradiação ao braço esquerdo e mandíbula, associada a sudorese e náuseas, com início, sem alívio pelo repouso.							1

	relieved by rest.									
--	-------------------	--	--	--	--	--	--	--	--	--

x. “Textos Mais Frequentes” – Avaliação da tradução do Google Tradutor

	ORIGINAL_EN	PT (Google)	Gr a	Ter	Po	Om	Tra	Mai	Coe
1	- Accidents and safety:.	- Acidentes e segurança:.							
2	- Adequate/inadequate feeding.	- Alimentação adequada / inadequada.							
3	- Bottle weaning:.	- Desmame Bottle :.					3		
4	- Child's development and temperament:	- Criança de desenvolvimento e temperamento:	3						
5	- Child's development, temperament and relationship:	- Desenvolvimento, o temperamento da criança e do relacionamento:	3						
6	- Child's reaction to strangers:	- Reação da criança a estranhos:							
7	- Cognitive development, laterality:.	- O desenvolvimento cognitivo, lateralidade:.				1			
8	- Cognitive development:.	- O desenvolvimento cognitivo:.				1			
9	- Complete physical examination was performed and the child's state is normal.	- Exame físico completo foi realizado e do estado da criança é normal.	1						
10	- Dentition and oral hygiene:.	- Dentição e higiene oral:.							
11	- Dentition and teeth brushing:	- Dentição e escovar os dentes:							
12	- Dentition and teeth brushing:.	- Dentição e escovar os dentes:.							
13	- Dentition, teeth brushing and use of dental floss:.	- Dentição, escovar os dentes e uso de fio dental:.							

14	- Dentition:	- Dentição:							
15	- Dentition:.	- Dentição:.							
16	- Discovery of sexual orientation:.	- Descoberta de orientação sexual:.							
17	- Eating habits:.	- Hábitos alimentares:.							
18	- Gait:.	- Marcha:.							
19	- General aspects:.	- Aspectos gerais:.							
20	- Harmful consumption:.	- Consumo Nocivo:.						1	
21	- Hearing/Language:	- Audição / Linguagem:							
22	- Hearing:	- Audição:							
23	- Issues related to accidents and security were/were not discussed.	- Questões relacionadas com acidentes de trabalho e de segurança foram / não foram discutidos.							
24	- Issues related to the child's nutrition were/were not discussed:.	- Questões relacionadas com a nutrição da criança foram / não foram discutidos:.	1						
25	- Issues related to the introduction of new food into a child's diet were/were not discussed.	- Questões relacionadas com a introdução de novos alimentos na dieta de uma criança foram / não foram discutidos.	1						
26	- Language development:	- O desenvolvimento da linguagem:				1			
27	- Language:	- Idioma:							

28	- Language:.	- Idioma:.							
29	- Learning of daily routines and rules:	- Aprendizagem de rotinas diárias e regras:							
30	- Lipid profile:.	- Perfil lipídico:.							
31	- Main diet errors:	- Principais erros de dieta:							
32	- New skills and acquisitions (motor and language):	- Novas competências e aquisições (motoras e de linguagem):							
33	- Normal development for age.	- O desenvolvimento normal para a idade.				1			
34	- Normal/abnormal hearing.	- A audição normal / anormal.				1			
35	- Normal/abnormal vision.	- Visão normal / anormal.							
36	- Oral health:.	- Saúde Oral:.							
37	- Orthopedic problems:	- Problemas ortopédicos:							
38	- Pacifier weaning:.	- Chupeta desmame:.	2						
39	- Participates/does not participate in sports and cultural activities.	- Participa / não participa de atividades esportivas e culturais.							
40	- Personality assertion, stubbornness, negativism.	- Afirmação de personalidade, teimosia, negativismo.							
41	- Physical exercise, outdoor activities, leisure and intellectual stimulation:.	- O exercício físico, actividades ao ar livre, lazer e estímulo intelectual:.				1			1

42	- Playing, drawing, television habits, going to bed ritual:	- Jogar, desenho , os hábitos de televisão, ir para a cama ritual :	2			1	1		
43	- Preparation for admission to school, prevention of school failure, attitude towards school and studying:.	- Preparação para a admissão à escola, a prevenção do fracasso escolar, a atitude em relação à escola e estudar:.				1			
44	- Psychomotor development.	- Desenvolvimento Psicomotor .						1	
45	- Psychomotor development:.	- Desenvolvimento Psicomotor :.						1	
46	- Pubertal stage:.	- Pubertal etapa:.	1	2					
47	- Puberty:.	- Puberdade:.							
48	- Reading habits:.	- Hábitos de leitura:.							
49	- Relationship with other children:.	- Relação com outras crianças:.							
50	- Sleeping/resting patterns, nocturnal enuresis:.	- Padrões de sono / descanso, enurese noturna:.							
51	- Sphincter control:	- O controle esfinteriano:				1			
52	- Temperament - negativism, stubbornness, jealousy, rivalry:	- Temperamento - negativismo, teimosia, ciúme, rivalidade:							
53	- Temperament and relationship - independence, separation anxiety, night terrors:	- Temperamento e relacionamento - independência, ansiedade de separação, os terrores noturnos:				1			
54	- Toys:	- Brinquedos:							

55	- Vaccination:.	- Vacinação:.							
56	- Vision:.	- Visão:.							
57	Fr nasogastric tube, for continuous drainage to collection bag with clamp / feeding by...	Pe. sonda nasogástrica, para drenagem contínua de bolsa coletora com braçadeira / alimentação por ...	2				2		
58	Fr urinary catheter, for continuous drainage to collection bag with clamp.	Cateter urinário Fr, para drenagem contínua a bolsa coletora com grampo.	1						1
59	Cardiotocography (CTG)	Cardiotocografia (CTG)							
60	Discharge conditions:	Condições de descarga:		3					
61	Admission date:	Data de admissão:							
62	Discharge date:	Data de descarga:		3					
63	Admission diagnosis:	Diagnóstico de admissão:							
64	Discharge diagnosis:	Descarga diagnóstico:		3					
65	... weeks pregnant.	... Semanas de gravidez.							1

66	Reason for visit:	Motivo da visita:							
67	Para	Para							
68	- Leisure occupations:.	- Lazer: ocupações.	1		3				
69	- Sleeping/resting patterns:.	- Padrões de sono / descanso:.							
70	- Social and cultural interests:.	- Interesses sociais e culturais:.							
71	- Statural and ponderal growth:.	- Estatural eo crescimento ponderal:.	2			1			
72	weeks of high-risk pregnancy.	 semanas de gravidez de alto risco.							
73	weeks of pregnancy without complications.	 semanas de gravidez sem complicações.							
74	? year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	? paciente anos de idade apresentando desconforto retroesternal aos esforços (fazer a cama, subir escadas, lavar pratos) e aliviada pelo repouso.	2						
75	? year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	? anos de idade do paciente, tabagista, com história familiar de hipertensão, com desconforto retroesternal grave que irradia para o braço esquerdo e mandíbula associada à falta de ar e náuseas, sem alívio com o repouso.	2						
76	... year old patient presenting retrosternal discomfort on exertion (making the bed, climbing stairs, washing dishes), and relieved by rest.	Ano de idade ... paciente com desconforto retroesternal aos esforços (fazer a cama, subir escadas, lavar pratos) e aliviada pelo repouso.	2						
77	... year old patient, smoker, with family history of hypertension, with severe retrosternal discomfort radiating to the left arm and jaw associated with shortness of breath and nausea, without relief by rest.	... Anos de idade do paciente, tabagista, com história familiar de hipertensão, com desconforto retroesternal grave que irradia para o braço esquerdo e mandíbula associada à falta de ar e náuseas, sem alívio com o repouso.	2						

78	weeks of high-risk pregnancy.	 Semanas de gravidez de alto risco.							1	
79	weeks of pregnancy without complications.	 Semanas de gravidez sem complicações.							1	
80	48 hours.	48 horas.								
81	Abdominal pain since...	Dor abdominal desde ...								
82	Abnormal genitourinary exam - Female.	Exame genitourinário anormal - Feminino.								
83	Abnormal genitourinary exam - Male.	Exame genitourinário anormal - Masculino.								
84	Abnormal haematologic system.	Sistema hematológico anormal.		1						
85	Absence / Temporary absence from clinical service due to.....	Ausência / ausência temporária do serviço clínico, devido à						1		
86	Admitted for: labs/antibiotics/therapies	Admitido para: laboratórios / antibióticos / terapias		2						
87	Foetal heart rate (FHR):	Frequência cardíaca fetal (FCF):								
88	Gravida; ...	gravida; ...		3						
89	Number of appointments.....	Número de consultas								
90	Number of appointments.....	Número de consultas								
91	Procedures executed in clinic:	Procedimentos realizados na clínica:								
92	Procedures executed in hospital:	Procedimentos realizados no hospital:								

93	Procedures executed in primary care center:	Procedimentos executados no centro da atenção primária:							
94	year old patient, smoker, with a history of arterial hypertension, has retrosternal oppressive pain with radiation to the left arm and mandible, associated with sudoresis and nausea, onset date/time , not relieved by rest.	anos de idade do paciente, fumante, com história de hipertensão arterial, tem dor opressiva retroesternal com irradiação para o braço esquerdo e mandíbula, associada a sudorese e náuseas, data de início / hora, não aliviada pelo repouso.	2						1

xi. Relatório de análise do SDL Trados Studio relativo aos textos “Instruções de Alta” e “Textos Mais Frequentes”

Analyze Files Report

Summary

Task: Analyze Files
Project: Mestrado
Translation Providers: EN_PT_ALLALL.sdltm
 Portuguese
 (Portugal)
Language:
Files: 2
Created At: 7/30/14 16:47
Task Duration: 3 seconds

Settings

Report Cross-file Repetitions: Yes
Report Internal Fuzzy Match Leverage: No
Report Locked Segments: No
Minimum Match Value: 65%
 Use best matches from all translation sources.
Search Mode:
Missing Formatting Penalty: 1%
Different Formatting Penalty: 1%
Multiple Translations Penalty: 1%
Auto-localization Penalty: 0%
Text Replacement Penalty: 0%

Totals

Total	Type	Segment s	Word s	Character s	Perce nt	Recognized Tokens	Tag s
Files:2	PerfectMatch	0	0	0	0.00%	0	0
Chars/Word:5.28	Context Match	2	2	12	0.12%	0	0
	Repetitions	0	0	0	0.00%	0	0
	Cross-file Repetitions	0	0	0	0.00%	0	0
	100%	35	188	1126	11.00%	2	0
	95% - 99%	24	180	1077	10.53%	0	0
	85% - 94%	21	142	912	8.31%	1	0
	75% - 84%	16	65	345	3.80%	0	0
	50% - 74%	4	8	40	0.47%	0	0
	New	94	1124	5505	65.77%	37	0
	Total	196	1709	9017	100%	40	0

File Details

File	Type	Segment s	Word s	Character s	Percen t	Recognized Tokens	Tag s
DischargeInstructions_EN.xlsx.sdlxliff	PerfectMatch	0	0	0	0.00%	0	0
Chars/Word:4.89	Context Match	1	1	6	0.09%	0	0
	Repetitions	0	0	0	0.00%	0	0
	Cross-file Repetitions	0	0	0	0.00%	0	0
	100%	2	4	37	0.35%	0	0
	95% - 99%	1	1	5	0.09%	0	0
	85% - 94%	0	0	0	0.00%	0	0
	75% - 84%	4	26	114	2.26%	0	0
	50% - 74%	0	0	0	0.00%	0	0
	New	93	1119	5467	97.22%	37	0
	Total	101	1151	5629	100%	37	0
SampleText_EN.xlsx.sdlxliff	PerfectMatch	0	0	0	0.00%	0	0
Chars/Word:6.07	Context Match	1	1	6	0.18%	0	0
	Repetitions	0	0	0	0.00%	0	0
	Cross-file Repetitions	0	0	0	0.00%	0	0
	100%	33	184	1089	32.97%	2	0
	95% - 99%	23	179	1072	32.08%	0	0
	85% - 94%	21	142	912	25.45%	1	0
	75% - 84%	12	39	231	6.99%	0	0
	50% - 74%	4	8	40	1.43%	0	0
	New	1	5	38	0.90%	0	0
	Total	95	558	3388	100%	3	0