

UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS



**Italiano Controlado para a Tradução
Automática:**

Italiano → Português

Giorgia Barberis

Tese orientada pela Prof.^a Doutora Palmira Marrafa,
especialmente elaborada para a obtenção do grau de
Mestre em Tradução

2017

AGRADECIMENTOS

Agradeço à Prof.^a Doutora Palmira Marrafa, pela orientação científica e pelo encorajamento nos momentos mais difíceis deste percurso.

Agradeço aos meus pais, por terem acreditado sempre em mim e por me terem apoiado moralmente e economicamente nesta experiência no estrangeiro.

Agradeço aos meus amigos e colegas, pelo apoio e pelo tempo dedicado.

ÍNDICE

Resumo	3
Abstract	4
1. Introdução	5
1.1 Objetivos do estudo	5
1.2 Metodologia	6
1.3 Organização do trabalho	8
2. A Tradução Automática.....	9
2.1 Breve história da tradução automática.....	10
2.2 Os diferentes tipos de sistemas de tradução automática	12
2.2.1 Sistemas de tradução automática orientados para o conhecimento.....	14
2.2.2 Sistemas de tradução automática orientados para os dados	16
2.2.3 Sistemas de tradução automática híbridos.....	18
2.3 O sistema Systran – SystraNet.....	19
2.4 Os problemas da tradução automática	22
3. As Linguagens Controladas.....	26
3.1 Breve história das linguagens controladas.....	26
3.2 Tipologia de linguagens controladas	28
3.3 Virtualidades e limites das linguagens controladas	30
3.4 Introdução à linguagem controlada para tradução automática do par italiano- português.....	31
4. Modo	37
4.1 Frases imperativas.....	38
4.2 Conjuntivo em frases simples	45
4.3 Frases Completivas	47
4.4 Frases Temporais	51

4.5 Frases Condicionais	55
4.6 Frases Concessivas	58
4.7 Conclusão.....	60
5. Modalidade.....	62
5.1 Verbos modais	62
5.1.1 <i>Dovere</i> > <i>Dever</i>	62
5.1.2 <i>Potere</i> > <i>Poder</i>	67
5.2 Expressões modais epistémicas	69
5.3 Conclusão.....	74
6. Aspeto	75
6.1 <i>Passato prossimo</i> e <i>Passato remoto</i>	75
6.2 Construções perifrásticas progressivas	80
6.3 Conclusão.....	86
7. Outros Fenómenos a requerer controlo.....	88
7.1 Aspetos Lexicais	88
7.2 Aspetos sintáticos	107
7.3 Conclusão.....	113
8. Conclusões e considerações finais	114
Anexo	116
Referências Bibliográficas.....	126
Referências Sitográficas	131

RESUMO

A importância da Tradução Automática (TA) prende-se com a necessidade de rapidez na comunicação, a qual está em constante evolução. Nas últimas décadas, de facto, a tradução automática tornou-se uma ferramenta imprescindível para um número cada vez maior de pessoas, visando a investigação neste âmbito a criação de sistemas de tradução automática de elevado desempenho, tanto em termos de eficiência como de qualidade. Apesar disso, posto que a máquina não pode substituir de forma totalmente satisfatória a tarefa do ser humano nesta matéria, os resultados da tradução podem apresentar inadequação ou agramaticalidade. Contudo, através da utilização das Linguagens Controladas (LC), ou seja da aplicação de um conjunto de restrições linguísticas aos textos de *input*, é possível obter resultados de melhor qualidade no texto de *output*.

O presente trabalho tem como objetivo estabelecer um conjunto de restrições sintáticas bem definidas e apresentadas sob a forma de regras declarativas para o par linguístico italiano-português, sendo o italiano a língua de *input* e o português a língua de *output*. Desta maneira, são estabelecidas restrições aplicáveis à língua italiana que permitem a obtenção de melhores resultados na tradução para o português. Os tópicos linguísticos objeto de análise são o modo, a modalidade e o aspeto. O sistema de tradução automática de referência é o SystraNet, disponível gratuitamente *online*. A linguagem controlada elaborada no presente trabalho é de tipo MOCL (*Machine-Oriented Controlled Language*), sendo que a alteração dos elementos linguísticos depende exclusivamente da aceitabilidade do resultado de tradução.

Palavras-chave: tradução automática, linguagem controlada, qualidade dos *outputs*.

ABSTRACT

The importance of Machine Translation (MT) rooted in the needs and constant evolution of rapid communication. During the last decade machine translation has become an essential tool for many people, as its main purpose has been aimed at the creation of automatic, high quality and high performance MT systems. Nevertheless, since a machine cannot fully replace the task of human beings, translation results can be inadequate or grammatically ill-formed. Through the application of Controlled Languages (CL), the usage of a set of language restrictions applied to natural languages, better quality results in the output text are obtained.

This study aims at establishing a set of well defined syntax restrictions presented in the form of declarative rules for the Italian-Portuguese language pair, where Italian is the input language and Portuguese is the output language. When applied to the Italian language, these restrictions determine the improvement of translation results in Portuguese. The linguistic issues analysed are mood, modality and aspect. The machine translation system used is SystraNet, available online for free. The controlled language elaborated on is classified as MOCL (Machine-Oriented Controlled Language), as the modification of the linguistic elements depends exclusively on the acceptability of the translation result.

Keywords: machine translation, controlled language, outputs quality.

1. INTRODUÇÃO

1.1 OBJETIVOS DO ESTUDO

Este trabalho tem como objetivo fundamental o estabelecimento de um conjunto de regras para o controlo de *inputs* em italiano a traduzir para português, com base na avaliação de *outputs* do sistema de tradução automática SystraNet.

Hoje em dia, os sistemas de tradução automática (TA) representam uma ferramenta de grande utilidade seja para o tradutor seja para o público em geral. No caso em que a tradução se destina apenas ao entendimento geral do conteúdo, a simples submissão de um dado texto a um sistema de tradução automática pode ser suficiente, embora o resultado nem sempre seja totalmente compreensível. Quando a tradução se destina à publicação de um determinado tipo de texto, contudo, a utilização da tradução automática deve ser combinada com a pré- e a pós-edição, para garantir a adequação dos *outputs*. O presente trabalho centra-se na criação de uma linguagem controlada a aplicar ao texto antes da sua submissão a um sistema de tradução automática, ou seja, na pré-edição.

A aplicação de regras de controlo da linguagem a certos tipos de *inputs* a serem submetidos à tradução automática contribui para a melhoria dos resultados de tradução, assim como para a eficiência do desempenho do sistema na execução da mesma (podendo as regras servir para uma ou mais línguas de *output*).

As linguagens controladas são formadas por restrições bem definidas e apresentadas sob a forma de regras declarativas, que têm, entre outros objetivos, o de eliminar, sempre que possível, as ambiguidades. A redação de textos de forma clara e coerente, com uma adequada eliminação das ambiguidades, facilita a análise dos textos pelos sistemas de tradução automática e permite a geração de *outputs* de melhor qualidade.

A razão desta escolha deve-se ao facto de o par de línguas italiano-português não ser muito trabalhado no âmbito das linguagens controladas. A não existência de uma linguagem controlada para a tradução automática de italiano para português confere, naturalmente, motivação a este trabalho.

O objetivo geral deste projeto é, portanto, permitir que a tradução para português seja mais adequada quando a língua de partida é o italiano. A utilização do italiano controlado e, conseqüentemente, a observação das regras estabelecidas, permite a obtenção de resultados de melhor qualidade com a conseqüente redução da intervenção humana numa eventual fase de pós-edição.

Os três grandes tópicos linguísticos aqui levados em conta são modo, modalidade e aspeto. Não cabe no âmbito de um trabalho desta natureza a identificação e o tratamento exaustivo de todos os fenómenos que durante o estudo levado a cabo evidenciam necessidade de controlo. Ainda assim, a partir dos casos estudados, foram selecionados e analisados, além dos fenómenos referidos, os aspetos linguísticos que requerem um maior desafio para a tradução automática e que, conseqüentemente, carecem igualmente de controlo.

1.2 METODOLOGIA

Criar e organizar o conjunto de regras para a linguagem controlada supõe o levantamento e a análise dos problemas para, subsequentemente, se elaborarem regras a testar no sistema de tradução automática em questão (SystraNet, neste caso). Depois da análise dos resultados dos *outputs* e a identificação das falhas da tradução, experimentaram-se diferentes estratégias para remover o problema e escolheu-se a mais adequada.

Os textos utilizados para os testes e para o desenvolvimento do *corpus* deste trabalho foram recolhidos nos vários livros e artigos indicados na bibliografia e outros foram adaptados ou construídos com recurso à introspeção, conforme o fenómeno linguístico em questão. Para sustentar a descrição dos vários fenómenos linguísticos foram utilizadas algumas gramáticas¹, bem como bibliografia especializada sobre temas específicos (modo, modalidade e aspeto).

Contudo, posto que a maioria dos exemplos utilizados para os testes são descontextualizados, a análise dos textos é feita com base nos elementos relevantes para a análise dos três grandes tópicos linguísticos centrais que constituem o objeto de estudo deste trabalho.

¹ Em particular, *Gramática da Língua Portuguesa*, (várias edições) de Maria Helena Mira Mateus, Ana Maria Brito, Inês Duarte, Isabel Hub Faria *et al.*, *Grammatica Italiana* de Serianni e *Grammatica italiana con nozioni di linguistica* (terceira edição) de Dardano e Trifone.

Em suma, o método de análise e desenvolvimento do *corpus* do trabalho envolveu os passos seguintes:

1. Observação de dados, análise e descrição das especificidades das línguas em contraste, à luz da bibliografia consultada.
2. Recolha de frases na língua italiana utilizáveis como exemplos representativos do fenómeno em questão para verificar se o sistema reconhece ou não certas estruturas e se as traduz de maneira correta.
3. Análise das problemáticas que podem apresentar os *outputs* da TA e procura dos fatores que causam as falhas (por exemplo, ambiguidades lexicais e estruturais).
4. Edição/alteração do *input*, em função do resultado que se quer obter na língua de *output*.
5. Avaliação das opções de edição do *input* em função da adequação do resultado de tradução e da validade do método utilizado (cuja determinação envolve, geralmente, sucessivas tentativas), aplicando-o a outros exemplos com problemas parecidos.
6. Criação de regras para a linguagem controlada.

Cabe lembrar que a estratégia adotada nesta análise visa à melhoria do *output* para que este seja gramaticalmente correto e para que o conteúdo da tradução corresponda ao da língua de partida. Para cumprir este objetivo, a qualidade do *input* não é relevante, pelo que não existem restrições no que respeita a tornar a frase de partida agramatical, sempre que tal se revele necessário para que o *output* resulte adequado. Em consequência, no âmbito deste trabalho, será permitido o uso de frases agramaticais em italiano com o objetivo de obter os resultados desejados em português.

Além da análise de questões de modo, modalidade e aspeto, a última parte do trabalho (capítulo 7) é dedicada a outros fenómenos encontrados nos exemplos utilizados. Estes fenómenos causam problemas de outra ordem à tradução automática e precisam, portanto, de uma segunda fase de análise e consequente controlo. Quando possível, são criadas regras para evitar a ocorrência de tais fenómenos.

1.3 ORGANIZAÇÃO DO TRABALHO

O trabalho está organizado em oito capítulos, o primeiro dos quais apresenta uma visão geral do trabalho e descreve a metodologia utilizada.

O segundo capítulo trata do tema da tradução automática. Neste capítulo é apresentada uma panorâmica histórica dos sistemas de tradução automática e são descritas as diferenças entre os diversos tipos de paradigmas. Uma secção é dedicada ao sistema de TA Systran e ao seu funcionamento na versão *online* (SystraNet). Também são tratadas as problemáticas mais comuns da tradução automática e apresentados alguns exemplos preliminares.

O terceiro capítulo trata das linguagens controladas. Após a apresentação de uma breve história, fala-se das tipologias de linguagens controladas e, em seguida, apresentam-se as suas virtualidades e limitações. Além disso, são introduzidas algumas regras básicas e a aplicação destas à linguagem controlada para a tradução automática do par italiano-português.

Desde o quarto até ao sexto capítulo são analisados os três grandes temas deste trabalho: modo, modalidade e aspeto. Em cada capítulo é efetuado um estudo de casos que respeita o esquema apresentado na secção 1.2. As regras para a linguagem controlada de todos os temas tratados são expostas em anexo e seguem a estrutura das de Marrafa, Amaro, Mendes e Ibrahim (2011).

O sétimo capítulo, como já anteriormente mencionado, retoma os fenómenos problemáticos identificados nos vários exemplos analisados com vista ao estabelecimento de regras para o controlo dos *inputs* em matéria de modo, modalidade e aspeto.

Finalmente, o oitavo e último capítulo é dedicado às conclusões e considerações finais, considerando as regras criadas para a linguagem controlada, como resultado do estudo, e possíveis desenvolvimentos futuros.

2. A TRADUÇÃO AUTOMÁTICA

A tradução automática pode ser definida como a tradução de um texto numa língua natural para outra através de um computador. O objetivo é transferir, através de software apropriado, o significado da língua original para a língua de chegada de forma rápida, sem má formação sintática, independentemente da extensão do texto.

Para tal, um sistema de tradução automática tem de reconhecer todos os elementos linguísticos e encontrar as devidas correspondências na outra língua. Este processo é determinado por mecanismos internos instalados no sistema, os quais permitem a análise de um dado texto na língua de partida e a geração do texto correspondente na língua de chegada.

A ideia de usar máquinas para traduzir textos é muito antiga, pois remonta aos séculos XVI e XVII, períodos nos quais havia uma grande confiança no poder da matemática e na sua aplicação em todos os campos. Todavia, a partir da criação das primeiras máquinas nos anos 50 do século XX, o investimento na área da tradução automática foi pouco regular, pois houve muitas falhas que impediram o sucesso da investigação.

Existem três tipos de necessidades informacionais que a tradução automática pode satisfazer: a disseminação, a assimilação e o intercâmbio (BANJAR, 2001: 26). A tradução com finalidades de disseminação começou a ser efetuada pelos sistemas de tradução automática a partir dos anos 60 do século XX. Contudo, para este fim e para a publicação satisfazer a qualidade desejada, o *input* precisa de ser controlado e o *output* precisa de uma revisão de um tradutor humano. A tradução com finalidades de assimilação serve para aqueles utilizadores que, mesmo tendo consciência da baixa qualidade da tradução, querem ter uma compreensão da informação geral do texto. Neste sentido, a tradução automática é destinada apenas à compreensão do texto, mas não à sua publicação. Também existe a tradução com finalidades de intercâmbio, ou seja, a tradução em tempo real de páginas web ou diálogos de *chat*. A investigação neste domínio está em progresso e a tecnologia em desenvolvimento contínuo, pelo que vem ganhando terreno a perspectiva de combinação de reconhecimento de voz e tradução automática para este fim.

Neste capítulo fala-se sobre a história da tradução automática, dos seus progressos ao longo dos anos e dos desafios que se têm vindo a colocar. Em seguida, são apresentados os diferentes tipos de sistemas de tradução automática, bem como o sistema utilizado no presente trabalho, o SystraNet. Por fim, são descritos os problemas mais comuns que se podem encontrar na tradução automática, para fornecer uma ideia prática de como lidar com eles e ultrapassá-los.

2.1 BREVE HISTÓRIA DA TRADUÇÃO AUTOMÁTICA

Como refere Hutchins (2003), a história da tradução automática tem início no princípio dos anos 30, quando apareceram os primeiros dicionários eletrónicos concebidos pelo engenheiro francês Georges Artstruni e pelo russo Petr Trojanskij. Estes dicionários tiveram bastante sucesso, mas não o suficiente para se apostar na investigação. Contudo, graças a eles, foi lançada a ideia da tradução automatizada através do computador, que permitiu o seu desenvolvimento nos anos 50.

Outro estímulo para a investigação na área da tradução automática foi dado pelo matemático americano Warren Weaver, o qual, especializado quer em linguística quer em informática, tentou divulgar a importância dos computadores para traduzir automaticamente. Apesar de ter sido desencorajado pelo MIT (*Massachusetts Institute of Technology*), Weaver teve um papel fundamental porque foi o primeiro que quis realmente apostar na tradução automática, sendo que tinha a convicção de que esta constituía o futuro da tradução.

Finalmente chegam os anos 50, década em que se começou a investigar mais na área da tradução automática. O investigador Yehoshua Bar-Hillel, do MIT, declarou logo no início da sua investigação que a tradução automática não podia obter sucesso sem ser combinada com a intervenção humana, tendo como obstáculo linguístico maior a ambiguidade. Em 1952, durante a primeira Conferência de Tradução Automática, falou da possibilidade de efetuar um controlo do léxico para diminuir os problemas semânticos. Durante aquela conferência foram propostas várias ideias, entre as quais um possível papel dos editores humanos durante a pré- e a pós-edição, ponto de partida muito importante para o estudo sobre as linguagens controladas. Em consequência, a tradução automática começou a ser considerada como um auxílio para o tradutor humano e o interesse neste campo de investigação começou a evoluir.

Nos anos 60, a questão tornou-se complicada, sendo que as traduções experimentadas nas máquinas ainda eram de baixa qualidade e havia muita dificuldade em resolver os problemas. Por conseguinte, o Estado americano começou a desinvestir na investigação, pelas razões apontadas no Relatório ALPAC (*Automatic Language Processing Advisory Committee*) do ano 1966. Esta comissão tinha sido criada com o fim de avaliar os progressos da tradução automática, que este relatório revelou serem demasiado fracos para o investimento até então realizado, revelando-se a investigação demasiado cara e sem futuro. Além disso, pensava-se que havia um número suficiente de tradutores humanos para satisfazer todas as necessidades, razão pela qual foram interrompidos muitos projetos que se estavam a desenvolver em várias partes do mundo.

Apesar de não haver, naquela altura, um particular interesse na área da sintaxe e da semântica por parte dos Estados Unidos, houve ainda assim investigadores que se ocuparam destas áreas com o fim de desenvolver abordagens diferentes.

Durante os anos 70, apesar do desânimo que causou o Relatório ALPAC, a investigação não parou nem sequer nos Estados Unidos, onde naquela altura se investigava muito sobre o par linguístico russo-inglês por razões relacionadas com a Guerra Fria.

Entre outros, foi criado na Universidade de Montreal o sistema METEO, pela necessidade de traduções rápidas de inglês para francês, dado o governo canadiano ser bilingue. Este sistema foi finalizado em 1977 e servia, basicamente, para passar informações meteorológicas no país. Foi uma inovação de particular importância porque representou o primeiro sistema completamente funcional, capaz de processar traduções completas e sem precisar de intervenção humana nem na pré- e nem na pós-edição.

Também a Europa continuou a investir na investigação na área da tradução automática. De facto, em 1976 foi instalado o sistema Systran² na Comissão Europeia, útil para a tradução de vastos documentos em línguas diferentes por parte das instituições europeias.

Naquela altura, um facto de particular importância foi o surgir das primeiras multinacionais e comunidades multilingues, para as quais a tradução automática se tornou um recurso fundamental. De facto, no final da década dos 70, não havia um número suficiente de tradutores humanos, pelo que a tradução automática se impunha como uma necessidade.

² O Systran é o sistema de referência deste trabalho; será referido de forma aprofundada na secção 2.3.

Graças às várias inovações na área e ao surgir das multinacionais, nos anos 80 foram desenvolvidos diversos sistemas e financiados diversos projetos, entre os quais cabe mencionar METAL, MU, SUSY, ROSETTA e EUROTRA, entre outros. A estratégia que foi adotada para a criação da maioria destes projetos baseava-se em representações do conhecimento linguístico.

O verdadeiro ponto de viragem ocorreu nos anos 90, quando se começou a investigar sobre os sistemas baseados na abordagem estatística e em *corpora* de textos já traduzidos. Além disso, também começaram a aparecer as primeiras ferramentas de auxílio à tradução (CAT *tools*) tendo-se dado início igualmente à investigação sobre linguagens controladas.

Graças ao desenvolvimento dos computadores e da internet, nos anos 90 o que mais interessava ao público era um rápido acesso às informações da rede, sem se importar com a qualidade da tradução. Neste contexto, foram criados sistemas de tradução automática gratuitos *online*, como por exemplo, o Babelfish e o Google Tradutor.

A partir do ano 2000, as abordagens adotadas foram basicamente a estatística e algumas ainda baseadas em regras. Graças ao desenvolvimento de tais abordagens, muitos investigadores têm vindo a desenvolver sistemas híbridos, que combinam diferentes métodos no mesmo sistema.

Hoje em dia, a tradução automática é usada por utilizadores comuns e também por tradutores profissionais, pois já não é considerada como concorrente do tradutor, mas sim como um auxílio no trabalho da tradução. O público em geral usa habitualmente estas ferramentas gratuitas – entre as mais conhecidas está o Google Tradutor – para aceder rapidamente a todo tipo de informação.

2.2 OS DIFERENTES TIPOS DE SISTEMAS DE TRADUÇÃO AUTOMÁTICA

A maneira de classificar os sistemas de tradução automática pode seguir diferentes critérios. Basicamente, classificam-se segundo os diferentes paradigmas em que se enquadram e de acordo com a sua arquitetura interna.

Os paradigmas dos sistemas de tradução automática são classificados da seguinte maneira:

- Paradigmas orientados para o conhecimento;
- Paradigmas orientados para os dados;
- Paradigmas híbridos.

Primeiro, surgem os paradigmas orientados para o conhecimento e só depois de algumas décadas de investimento nesse sentido começa a mudar-se de paradigma, com o desenvolvimento de sistemas orientados para os dados. Por último, naturalmente, surgem os paradigmas híbridos.

Os sistemas híbridos foram criados com o objetivo de desenvolver diferentes estratégias de resolução conforme o tipo de problema de tradução a resolver. De facto, este tipo de sistemas é uma combinação de componentes de representação de conhecimento linguístico e componente estatística, podendo a hibridação ser, por sua vez, orientada para os sistemas baseados em regras ou para os sistemas baseados em *corpora*.

Por cada paradigma podem distinguir-se diferentes sistemas, tendo em conta a sua arquitectura e o seu modo de funcionamento. A arquitectura é a organização dos módulos de processamento do sistema e as abordagens diferenciam-se conforme os níveis de análise. Estas abordagens podem ser:

- Direta;
- Por transferência;
- Por interlíngua.

A abordagem de tradução direta respeita à “primeira geração” dos sistemas de tradução automática (desde os anos 50 até metade dos anos 70), enquanto as abordagens por transferência e por interlíngua (anos 70 e 80) são características da “segunda geração” (HUTCHINS, 2003: 504).

Esta classificação é ilustrada no triângulo de Vauquois, proposto pelo matemático francês Bernard Vauquois:

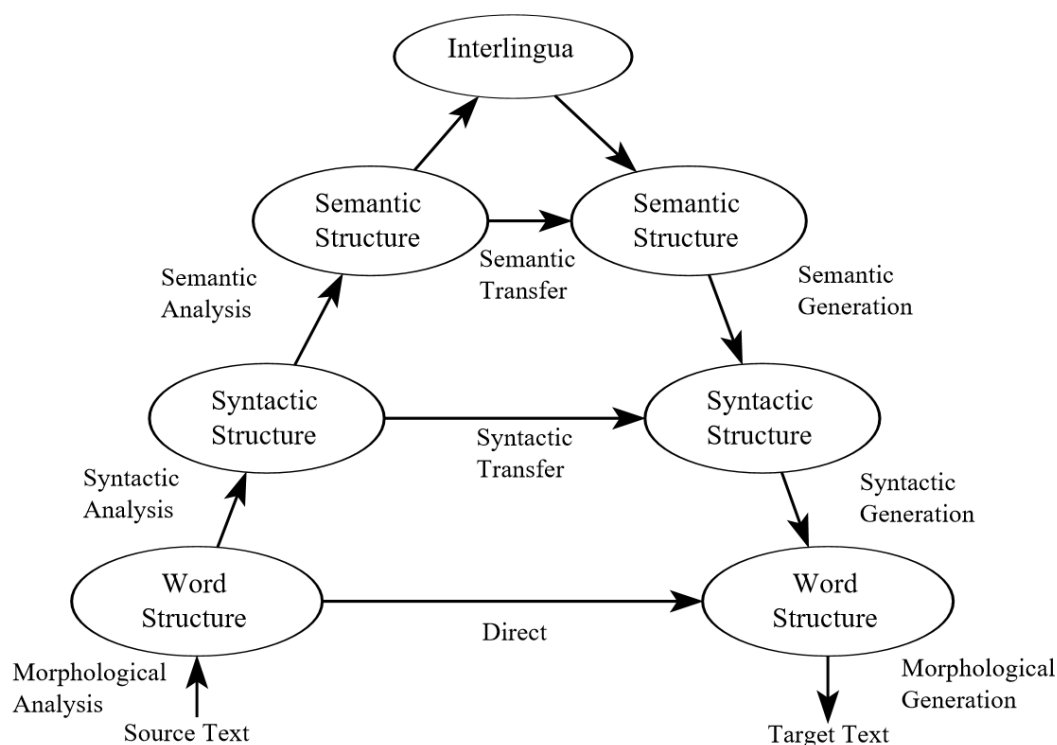


Figura 1 – Triângulo de Vauquois (extraído de DOOR, JORDAN, BENOIT, 1998: 13)

Os diferentes níveis representados no triângulo correspondem aos níveis de profundidade da análise e da geração. A tradução direta, situada na base do triângulo, é a tradução palavra a palavra com a respetiva análise/geração morfológica, sem nenhuma outra informação de ordem linguística. Entre a base e o vértice estão as abordagens por transferência sintática e semântica, cuja tradução se baseia em representações de tais estruturas no texto de partida. O topo do triângulo é constituído pela interlíngua, ou seja, uma representação conceptual (construída a partir da língua de partida), num formato independente das línguas.

2.2.1 SISTEMAS DE TRADUÇÃO AUTOMÁTICA ORIENTADOS PARA O CONHECIMENTO

Este tipo de sistemas de tradução automática foram os primeiros a serem criados e também os primeiros a despertarem o interesse dos investigadores. Apesar de haver discordâncias quanto à classificação dos diferentes sistemas de base linguística, todos estes têm em comum a integração de representação do conhecimento linguístico codificada sob a forma de regras.

Basicamente, estes sistemas centram-se no uso de representações do léxico, da sintaxe e da semântica para efeitos da análise da língua de partida e da geração da língua de chegada. De facto, o termo *knowledge* em *knowledge-based* designa o conhecimento linguístico nas suas diversas componentes, pois uma alta qualidade de tradução requer a completa compreensão do texto de partida (Arnold *et al.*, 1994 *apud* Banjar, 2001: 41), bem como o uso articulado das referidas componentes na geração do texto de chegada.

A abordagem mais antiga adotada para os sistemas orientados para o conhecimento baseia-se na tradução direta (como se pode observar no triângulo de Vauquois, figura 1). Neste tipo de sistema, a programação é estabelecida para apenas uma combinação linguística e a tradução processa-se diretamente do texto fonte para o texto alvo. O texto não é analisado a nível sintático mas é traduzido palavra a palavra, seguindo as regras morfológicas da língua de chegada. Os sistemas que se baseiam nesta abordagem consistem apenas num dicionário bilingue e num programa de análise e geração de textos.

Dentro do paradigma orientado para o conhecimento, os sistemas mais comuns são os baseados em regras (RBMT). Este tipo de sistemas pode seguir duas abordagens diferentes, baseadas na tradução por transferência ou por interlíngua³.

A abordagem por transferência, a mais largamente utilizada, envolve três fases: análise, transferência e geração. O texto da língua de partida é analisado e representado estrutural e semanticamente (fase 1), e subsequentemente, estas representações são convertidas em representações equivalentes na língua de chegada (fase 2), com a consequente geração do texto nesta língua (fase 3).

Esta abordagem é a mais utilizada hoje em dia e é adotada também pelo sistema Systran, entre outros.

A outra abordagem é de interlíngua, que envolve um nível de representação conceptual dos textos da língua de partida, independente das línguas, a partir do qual são gerados os textos correspondentes nas diferentes línguas de chegada. As representações na interlíngua são o ponto intermédio de execução da tradução, sendo que o texto passa por vários níveis de análise (morfológica, sintática e semântica) e por um nível de representação conceptual – interlíngua – a partir do qual se inicia o processo, inverso, de

³ Não sendo pacífica a inclusão da abordagem direta nos RBMT e dado que as regras envolvidas tanto na análise morfológica como na sua geração são muito rudimentares, neste trabalho situa-se esta abordagem fora dos RBMT.

geração do *output*. Esta abordagem é ideal para traduzir expressões complexas e para a tradução multilingue, sendo que a interlíngua é basicamente uma representação semântica da língua de partida através da qual se pode gerar um texto em qualquer língua.

Contudo, nos dois casos de tradução indireta (por transferência e interlíngua), o pressuposto comum é a necessidade de ter uma representação intermédia do significado da língua de *input* para gerar o seu equivalente na língua de *output*. A diferença principal entre as duas abordagens é o facto de a interlíngua ser uma representação independente das línguas, enquanto na abordagem por transferência as representações são todas específicas das línguas.

2.2.2 SISTEMAS DE TRADUÇÃO AUTOMÁTICA ORIENTADOS PARA OS DADOS

Contrariamente aos sistemas orientados para o conhecimento linguístico, estes sistemas não se baseiam nem em teorias nem em propriedades linguísticas, enquadrando-se, por conseguinte, num paradigma distinto. Existem dois tipos de sistemas de tradução automática orientados para os dados: os sistemas estatísticos (SMT) e os sistemas baseados em exemplos (EBMT). As abordagens adotadas por este tipo de sistemas dependem da existência de *corpora* de textos, que são basicamente grandes colecções de textos e respectivas traduções.

Como já mencionado acima, foram concebidos a partir dos anos 90 e o seu aparecimento constituiu uma grande inovação, bem como um grande desafio para a tradução automática, pois em muito pouco tempo passaram a dominar a área tornando-se cada vez mais populares.

Os sistemas estatísticos efetuam uma análise estatística de *corpora* paralelos bilingue, através do alinhamento de palavras, expressões e sequências de palavras operando por meio de cálculo de probabilidades. Alguns sistemas estatísticos incorporam uma língua ponte (ou língua Pivot) no sistema para facilitar a tradução, que permite criar uma “ponte” através das duas línguas (fonte e alvo) de modo a ultrapassar dentro do possível o carácter deficitário dos *corpora*.

Este modelo de sistemas de tradução automática não requer nenhum conhecimento preliminar de um dado par de línguas e, por conseguinte, não segue nenhuma regra linguística. Além do mais, o facto de não serem adaptados a apenas um par de línguas,

permite uma rápida adaptação a diferentes combinações linguísticas, tornando mais rápida a tradução automática de e para mais línguas. Por estas e outras razões, os sistemas estatísticos têm sido aqueles em que mais se tem investido, sobretudo durante a última década.

O Google Tradutor é um exemplo de sistema de tradução essencialmente estatístico. É uma ferramenta rápida e gratuita, utilizada pela grande maioria do público. No entanto, este sistema tem algumas limitações, devidas, fundamentalmente, à escassez de *corpora* em algumas línguas, pelo que, por vezes, tem que passar por uma outra língua (inglês) para poder executar a tradução.

Outro exemplo importante entre os sistemas orientados para os dados são os baseados em exemplos. Este modelo foi estudado a partir dos anos 80, mas foi desenvolvido só nos anos 90.

Este tipo de sistemas tem como objetivo construir o *output* com base na similitude de textos em diferentes línguas integrados na base de dados, reconhecendo a semelhança de uma frase ou expressão na língua de partida com uma frase ou expressão anteriormente traduzida na língua de chegada, através de uma “tradução por analogia” (Door, Jordan, Benoit, 1998: 33). Basicamente, o sistema efetua uma comparação do *input* com *corpora* de exemplos representativos já traduzidos e daí combina os dados conforme as correspondências mais próximas, formando uma nova frase na língua de *output* que será utilizada como modelo das traduções para aquela combinação linguística específica.

Mais tecnicamente, a tradução é executada em três fases: *matching* (correspondência), *alignment* (alinhamento) e *recombination* (recombinação). Durante o *matching*, o sistema procura os fragmentos de exemplos semelhantes na língua de partida num conjunto de textos monolíngues. Daí o sistema identifica a correspondência entre estes fragmentos e as traduções através do alinhamento de dados (*alignment*) e, por último, junta todos os fragmentos de exemplos encontrados, combina-os e reordena-os para poder construir a frase na língua de chegada (*recombination*).

Os sistemas baseados em exemplos distinguem-se dos sistemas estatísticos por fazerem um uso mais amplo de técnicas de análise estatísticas e por não serem limitados necessariamente a um *corpus* específico.

Quer os sistemas orientados para o conhecimento quer os sistemas orientados para os dados têm vantagens e desvantagens. No primeiro tipo de sistemas, a construção de regras linguísticas visa representar o conhecimento linguístico. O problema desta abordagem está no facto de essa representação envolver grande complexidade e requerer formalismos com custos de resolução elevados, que, além do mais, acarretam sérios prejuízos para a eficiência do sistema. Daí os orientados para os dados. A vantagem principal deste tipo de sistemas, quando comparados com os sistemas orientados para o conhecimento, é a qualidade das traduções dos *corpora*, pois são, de um modo geral, efetuadas por tradutores profissionais e, por conseguinte, há garantia de que os resultados dos *outputs* são característicos de falantes nativos. Por outro lado, dado estes sistemas serem completamente dependentes de *corpora*, a obtenção de um bom resultado depende da amplitude do mesmo *corpus*. Quanto mais limitado for o *corpus*, piores serão os resultados. Além disso, os sistemas estatísticos são limitados do ponto de vista da dependência contextual, enquanto os baseados em exemplos apresentam particulares dificuldades com o processamento de estruturas fráscas complexas.

2.2.3 SISTEMAS DE TRADUÇÃO AUTOMÁTICA HÍBRIDOS

Graças à introdução de novas aplicações para os processos de tradução automática, foi possível compreender os limites da utilização de apenas uma abordagem para os problemas de tradução. A partir da consciência de que os sistemas orientados para os dados não se revelaram capazes de ultrapassar todas as lacunas dos orientados para o conhecimento e de que as duas abordagens tinham diferentes tipos de vantagens, chegou-se à conclusão de que a combinação das duas abordagens poderia trazer um incremento na melhoria dos resultados. Como escreve Hutchins (2003: 17),

“[i]t is now widely recognised that there can be no single method for achieving good-quality automatic translation, and that future models will be ‘hybrids’, combining the best of rule-based, statistics-based and example-based methods.”

Existem vários tipos de sistemas híbridos. Citando Costa-Jussà e Fonollosa (2015: 4), “a hibridação pode ser orientada para os sistemas baseados em regras ou orientada para os sistemas baseados em *corpora*.”

Como explicado pelos mesmos autores, no caso da hibridação orientada para os sistemas baseados em regras, existem técnicas diferentes. Por exemplo, pode ser usado um *corpus* para construir um sistema baseado em regras, de modo a ampliar os dicionários com frases e exemplos retirados de *corpora* paralelos. Segundo estes autores, é também comum a técnica em que se faz uso de ferramentas baseadas em *corpora* para melhorar o *output* baseado em regras. Neste caso, o melhoramento do *output* é baseado em regras e ferramentas linguísticas, mas, antes do processo de transferência, usam-se excertos de traduções fornecidos pelos SMT, escolhendo as combinações mais prováveis entre os excertos. Cabe mencionar também os sistemas RBMT, que introduzem técnicas de aprendizagem automática para a identificação de opções de tradução apropriada. Entre estes últimos está o sistema Systran, o qual, com a construção de módulos de inferência estatística, demonstrou que os RBMT têm capacidade de aprendizagem a nível linguístico, tais como classificações morfológicas e informações semânticas e sintáticas do *corpus*.

No segundo tipo de hibridação, os sistemas podem incorporar regras e combinar abordagens baseadas em *corpora*. Alguns destes sistemas usam regras para reordenar a frase da língua de *input* o mais em consonância possível com a ordem da língua de *output*.

Enfim, pode haver também hibridações entre abordagens baseadas em *corpora*, mais geralmente entre abordagens estatísticas e abordagens baseadas em exemplos.

Na secção seguinte, fala-se do sistema Systran, sistema híbrido desde há poucos anos, utilizado no presente trabalho.

2.3 O SISTEMA SYSTRAN – SYSTRANET

O sistema SystraNet é um sistema híbrido porque combina conhecimento linguístico com uma componente estatística. Contudo, esta componente estatística foi introduzida só a partir do ano 2009, sendo este sistema inicialmente concebido como um sistema baseado em regras.

A génese deste sistema remonta aos primeiros anos de investigação na área da tradução automática, ou seja, aos anos 50. Teve origem a partir das experiências em tradução automática efetuadas na *Georgetown University* em 1954, com a contribuição da IBM e

foi um dos poucos sistemas de tradução automática a sobreviver aos cortes de financiamento depois do relatório ALPAC.

Foi criado por Peter Toma nos Estados Unidos, no ano de 1968, com a intenção de desenvolver um sistema de tradução automática para a USAF (*United States Air Force*) com uma única combinação linguística russo-inglês, por causa das necessidades da NASA durante a Guerra Fria. Inicialmente, a qualidade das traduções não era das melhores, mas o que mais importava então era obter traduções suficientemente compreensíveis para entender o conteúdo do texto de partida.

Em Junho de 1975, foi desenvolvida a combinação linguística inglês-francês para a CCE (Comissão das Comunidades Europeias). A partir daquele momento, o sistema mostrou a sua eficiência e, em consequência, foram criadas outras versões com novas combinações linguísticas, segundo as exigências da CCE. Nos anos seguintes, várias empresas começaram a investir neste sistema, que passou a ser utilizado para a tradução de documentos especializados em diferentes áreas técnicas, com dicionários específicos por cada área. Um momento culminante para o sistema é a *World Systran Conference*, em 1986, a primeira conferência dedicada a apenas um sistema. Nesta conferência foram discutidos desenvolvimentos futuros do sistema pelos seus principais utilizadores.

Entre 1986 e 1987 tornou-se disponível para o público geral e foi o primeiro sistema a conceber o *Customer Specific Dictionary*, em 1989, uma funcionalidade do sistema que permite a criação de dicionários pelos utilizadores, conforme as necessidades do cliente, permitindo, assim, uma melhor qualidade dos *outputs*.

Como já mencionado no início desta secção, 2009 foi um ano muito importante porque foi lançada a versão 7.0 do sistema, com a introdução da componente estatística.

Conforme explicado na página oficial da empresa⁴, a grande inovação deste mecanismo híbrido são as técnicas estatísticas, que permitem a aprendizagem automática a partir de dados (*corpora*) – monolíngue e multilíngue – e que, consequentemente, melhoram as fases do processo de tradução. Quem decide comprar o software terá também a possibilidade de personalizá-lo. Assim, o cliente tem a possibilidade de inserir na memória do sistema toda a terminologia e os recursos linguísticos para poder obter uma tradução de qualidade num dado domínio.

⁴ <http://www.systransoft.com.br/systran/perfil-corporativo/tecnologia/systran-hybrid-technology/>

A versão *online* gratuita da Systran é o sistema SystraNet e os resultados do seu desempenho são estudados no presente trabalho para a elaboração da linguagem controlada. Sendo a qualidade do *output* dependente do tamanho e da complexidade do texto de chegada, e não havendo na versão *online* maneira de personalizar os recursos linguísticos para a execução da tradução, o sistema não pode gerar um *output* de boa qualidade sem a intervenção humana. De facto, na secção de suporte do sistema, é possível encontrar algumas dicas para o utilizador melhorar a qualidade da tradução⁵, onde é especificado que o sistema não pode substituir de maneira nenhuma a tradução humana.

Dentre as dicas para a elaboração do texto *input* no sistema, cabe mencionar as mais importantes. Para que o sistema consiga executar corretamente a tradução para a língua de chegada, é necessário escrever frases com sujeito, verbo e estrutura frásica correta. O texto não deve ser demasiado longo. Sendo o caso, o utilizador terá de abreviá-lo ou dividi-lo em segmentos para efeitos da tradução. Além disso, é preciso usar o vocabulário apropriado, respeitar a gramática, a pontuação e os acentos e, sobretudo, evitar as expressões idiomáticas e dialetais, bem como os acrónimos.

Também há outra secção em que são dadas indicações específicas para o utilizador controlar o texto de partida, retomando as dicas acima mencionadas e salientando o facto de que “translation quality is determined by the quality of the text you translate”⁶. Estas indicações são listadas em seguida:

- ***Be Direct. Write in a Simple, Clear Manner.*** “Evitar estruturas complexas, frases de sentido ambíguo, orações extra, fragmentos de frase e palavras desnecessárias.”
- ***Be Concise and To The Point.*** “Usar frases curtas, mas não omitir qualquer constituinte da frase.”
- ***Do Not Leave Out Necessary Words.*** “A língua inglesa permite-nos claramente transmitir as nossas intenções, embora omitemos algumas palavras, como pronomes relativos (*who, whom, that, which*), preposições ou constituintes de verbos. Estas palavras são requeridas em outras línguas e devem ser incluídas nos documentos a serem traduzidos.”

⁵ <http://www.systranet.com/systranet-help/help-improve-translation-quality>

⁶ <http://www.systranet.com/it/systranet-help/help-writing-tips>

- ***Beware of Slang and Colloquialisms.*** “Evitar o uso de expressões idiomáticas nos documentos que se tenciona traduzir, pois estes termos variam de país para país. Os termos comumente usados em inglês não serão traduzidos corretamente para o uso em outras línguas.”
- ***Insert Proper Punctuation.*** “Assegurar-se de que se insira uma pontuação apropriada. A pontuação oferece orientação quer para humanos quer para computadores. Usa-se para dividir uma frase em partes lógicas. Sem uma correta pontuação, as frases podem ser interpretadas de várias formas.”
- ***Check for Accurate Spelling.*** “Verificar a ortografia e usar o corretor ortográfico antes de clicar no ícone da tradução. Se se der ao computador uma informação incorreta, a tradução pode ser interpretada mal.”
- ***Use Articles Whenever Possible.*** “Um artigo é uma palavra usada para indicar um nome e estabelecer a sua intenção. Por exemplo, em inglês, o artigo definido é *the* e os indefinidos são *a* e *an*. O uso de artigos reduz ambiguidades.”
- ***Consistent Use of Terminology and Abbreviations.*** “Usar sempre a mesma palavra, frase ou abreviação para descrever o mesmo objeto ou acção cada vez que estas aparecem num documento. Construções inconsistentes podem causar confusão quer para humanos quer para computadores.”
- ***Maintain a Simple Format.*** (Regra de formato não relevante para o presente trabalho)

Obviamente, todas as indicações para a redacção do texto de *input* são válidas para todos os pares linguísticos instalados no sistema e devem ser aplicadas pelos utilizadores cada vez que decidirem traduzir um texto com o SystraNet.

2.4 OS PROBLEMAS DA TRADUÇÃO AUTOMÁTICA

Como explicado na secção precedente, para que um sistema de tradução automática consiga executar o mais adequadamente possível a tradução, é preciso ter em conta os eventuais problemas que se colocam no texto de *input* e, se possível, resolvê-los previamente.

Tais problemas podem ser devidos a diferentes fatores: ambiguidade lexical e estrutural, não correspondência lexical ou gramatical, problemas de transferência lexical e estrutural e de reconhecimento das expressões idiomáticas.

A resolução da ambiguidade é certamente um dos maiores desafios para os investigadores da área da tradução automática e também uma questão muito tratada. A tradução executada por um tradutor humano precisa não só de conhecimentos linguísticos (gramática, semântica e pragmática), mas também de conhecimentos extralinguísticos (tipo de domínio, conhecimento do mundo e senso comum).

Como atrás referido, a ambiguidade pode ser de dois tipos: lexical e estrutural.

A ambiguidade lexical tem a ver com a multiplicidade conceptual e categorial de uma dada unidade lexical. Segundo Weinreich (1964, *apud* Pustejovsky, 1995: 27-28) a ambiguidade lexical é de dois tipos: ambiguidade complementar e contrastiva.

A ambiguidade complementar (ou polissemia complementar) “involves lexical senses which are manifestations of the same basic meaning of the word as it occurs in different contexts”. Um exemplo para este tipo de ambiguidade é a designação de *porta* como abertura e como objeto físico. Contudo, em grande parte dos casos, esta ambiguidade não causa problemas à tradução automática, pois o fenómeno tem idêntica distribuição em todas as línguas, pelo que não há ambiguidade ao nível de tradução.

A ambiguidade contrastiva é a que geralmente causa mais problemas ao sistema. Este tipo de ambiguidade “is seen where a lexical item accidentally carries two distinct and unrelated meanings (i.e. homonymy)” (Ibidem, pp. 27-28). Nestes casos, as unidades têm mais do que um sentido lexical e os conceitos não têm nenhuma relação entre eles. Um exemplo na língua italiana pode ser o substantivo *pesca*, que significa seja *pêssego*, seja *pesca*, palavra que causa problemas à tradução automática por ter mais do que uma correspondência na língua de chegada. Outro exemplo de ambiguidade contrastiva é o verbo *portare* (em português *levar* e *trazer*) o qual, conjugado na terceira pessoa do singular, na tradução automática pode ter três aceções diferentes: *traz*, *leva*, *porta* (substantivo). Contudo, pode haver casos em que o substantivo seja reconhecido pelo sistema, se for precedido de determinante. No caso da ambiguidade entre *trazer* e *levar*, o sistema não pode fazer uma distinção porque não tem informação sobre o contexto discursivo.

Em todos os casos de ambiguidade lexical, uma solução para a desambiguação pode ser escolher os equivalentes mais adequados conforme o contexto durante o processo de pré-edição (Hutchins, 2003: 502).

A ambiguidade estrutural ocorre quando a estrutura de uma frase pode ser analisada de mais do que uma maneira. Segundo Hutchins e Somers (1992: 88), as ambiguidades podem ser “reais” (nas quais o tradutor humano pode encontrar interpretações diferentes) ou próprias do sistema (as quais o tradutor humano nem sempre as reconhece como tais). As ambiguidades ditas “reais” precisam de um contexto específico para o tradutor humano as reconhecer, sem o qual não seria possível identificar o significado. Um exemplo deste tipo de ambiguidade pode ser a frase *L'uomo vide la ragazza con il telescopio*⁷, sendo que não fica claro se o telescópio é o instrumento com o qual o homem vê a rapariga ou se a rapariga é a que tem o telescópio.

As ambiguidades do sistema diferenciam-se das “reais” por terem a ver com a dificuldade do sistema no processo de tradução. Como afirmam os mesmos autores,

“[a]ccidental structural ambiguities occur due to an accidental combination of words having category ambiguities, due to alternative grammatical uses for syntactic constituents, or due to different possible combinations of syntactic constituents” (Hutchins, Somers, 1992: 89).

Um problema pode ter a ver, por exemplo, com a interpretação da co-referência, sobretudo nos casos das línguas tratadas no presente trabalho. O italiano e o português são línguas que permitem sujeito nulo, o que não acontece, por exemplo, em línguas como o francês. Tomando em consideração a frase *Maria ha detto a Giovanni che è stanca*⁸, a não realização do sujeito de *è stanca* não coloca problemas de interpretação aos falantes. Porém, o sistema decide pela co-referência com o sintagma nominal mais próximo, não estabelecendo qualquer relação sintática entre a parte não verbal do predicado – *stanca* – e o sujeito.

A questão da não correspondência também é considerada um desafio importante para os sistemas de tradução automática. Este problema também pode ser de tipo lexical ou estrutural. A não correspondência lexical deve-se ao facto de as línguas terem uma diferente categorização do mundo, pelo que uma determinada palavra pode existir numa língua e não existir uma correspondente noutra. A não correspondência gramatical pode ser devida ao facto de duas línguas terem diferentes estratégias linguísticas para expressar idênticos aspetos conceptuais ou terem estratégias idênticas para expressar

⁷ Tradução: “O homem viu a rapariga com o telescópio”, exemplo adaptado a *The man saw the girl with the telescope*, frase comunemente utilizada para ilustrar a ambiguidade estrutural.

⁸ Tradução: “Maria disse ao Giovanni que está cansada”, exemplo inventado. A solução para este tipo de problemas será tratada na secção 3.4.

aspectos conceptuais diferentes. Por exemplo, a existência do conjuntivo futuro ou do infinitivo pessoal em português e a inexistência destas formas verbais em italiano fazem com que o sistema não encontre uma correspondência e, por conseguinte, não traduza corretamente. Além disso, os problemas podem ser relacionados também com a ordem dos constituintes da frase ou com a omissão destes.

No capítulo seguinte são tratadas as linguagens controladas enquanto recursos de pré-edição de um texto a ser traduzido num sistema de tradução automática. São explicitadas as propostas de resolução dos problemas que acabamos de referir, com vista à obtenção de bons resultados na tradução automática.

3. AS LINGUAGENS CONTROLADAS

O conceito de linguagem natural controlada (LNC), ou linguagem controlada (LC), por abreviação, não tem uma definição uniforme na literatura, face à diversidade das abordagens e dos contextos de aplicação em que é utilizado. No âmbito deste trabalho, este conceito é entendido como um conjunto de restrições linguísticas a aplicar na escrita de textos numa dada língua natural.

As linguagens controladas podem ser de vários tipos e normalmente são classificadas conforme os seus objetivos. Podem ser aplicadas para várias finalidades, entre as quais:

- Melhorar a comunicação entre seres humanos;
- Melhorar a tradução automática.

A partir dos anos 30, as diferentes abordagens às linguagens controladas emergiram em diferentes disciplinas e em diferentes contextos (que não apenas o académico), tendo vindo, desde então, a ser objeto de investigação.

Neste capítulo, é apresentada uma breve história das linguagens controladas e uma classificação com base em diferentes critérios, com particular incidência no seu uso e modo de funcionamento na área da tradução automática.

Em seguida, fala-se das vantagens e das limitações das linguagens controladas e apresenta-se também uma parte introdutória da linguagem controlada para o par linguístico italiano-português.

3.1 BREVE HISTÓRIA DAS LINGUAGENS CONTROLADAS

A primeira ideia que levou à criação das linguagens controladas nasceu nos anos 30, quando um grupo de linguistas e estudantes sentiram a necessidade de criar uma variedade “base” de inglês para permitir ao maior número de pessoas o seu entendimento e uso em qualquer âmbito. Esta variedade “base” foi chamada *Basic English* e diferencia-se das outras tentativas de criação de linguagens universais por não ser uma língua híbrida ou criada artificialmente (como, por exemplo, o Esperanto). De facto, o *Basic English* pertence totalmente à língua inglesa, assentando na redução do número de palavras (de 75.000 para 850) e na simplificação das regras gramaticais, para

permitir a elaboração de qualquer tipo de texto de uma forma mais simples, utilizando determinados verbos, nomes e adjetivos⁹. Para a eventual elaboração de um texto técnico sobre uma determinada disciplina, pode-se adotar o *Basic English* com a complementação de um vocabulário técnico para a área em questão. Desta maneira, achava-se que, em vez de traduzir manuais técnicos ou textos desse género para todas as línguas que sejam necessárias, uma boa redação em *Basic English* facilitava a leitura e a compreensão por parte de engenheiros e mecânicos com conhecimentos limitados em inglês (Arnold *et al.*, 1994: 148).

Posteriormente, quando foram introduzidos os sistemas de tradução automática, o conceito de linguagem controlada foi desenvolvido com finalidades de tipo diferente. Como se pode ler em Banjar,

“[t]he concept of controlled language has been developed to form one of the principle of Computational Linguistics, particularly in the domain of Machine Translation. In order to get an acceptable translation, linguistic rules must be defined according to the ability of the machine translation system. These rules defined a controlled language. Adapting such a language to machine performances will allow good communication” (2001: 37).

Assim, foram criadas linguagens controladas para serem aplicadas à tradução automática. Desta maneira, reduziram-se os custos da tradução pelo facto de a máquina conseguir traduzir mais rapidamente e com melhor qualidade, pelo que a revisão podia resultar mais barata.

Graças à introdução das linguagens controladas, as empresas tiveram a possibilidade de inserir algumas delas nos seus próprios sistemas. De facto, era possível integrar dicionários terminológicos nos sistemas das empresas para a tradução de manuais técnicos para várias línguas, o que permitia também a criação de memórias de tradução e, com isso, a reutilização dos referidos recursos.

Entre as várias linguagens controladas criadas, além do já mencionado *Basic English*, cabe mencionar algumas das mais importantes e que tiveram mais sucesso.

O *Caterpillar Fundamental English*, por exemplo, foi criado pela empresa americana Caterpillar Inc. nos anos 70. Esta empresa é uma multinacional de fabricação de máquinas e motores, pelo que precisava da elaboração de manuais de carácter técnico

⁹ “This lexical economy was to be achieved in part by using ‘operator verbs’ with the set of nouns and adjectives to stand in for the vast number of derived verbs which are frequently used” (Arnold *et al.*, 1994: 147).

para serem traduzidos para várias línguas. Dada a expansão das áreas da empresa nas décadas seguintes e a consequente necessidade de ampliar o vocabulário, o *Caterpillar Fundamental English* tornou-se *Caterpillar Technical English*, com a introdução de novas entradas para uma maior cobertura e consistência a nível terminológico (Kapraht *et al.*, 1998 *apud* Zamagni, 2015: 54).

Outro exemplo de linguagem controlada é o *ASD Simplified Technical English*, (ASD-STE100). Foi criada pela empresa AECMA (*European Association of Aerospace Industries*) e inicialmente o projeto chamava-se *AECMA Simplified English*. Este projeto tinha o objetivo de investigar sobre a utilização de uma forma de inglês simplificada para a redação dos documentos pelos fabricantes de material aeroespacial. Teve particular sucesso pela sua flexibilidade e, por isso, foi aplicado também a contextos diferentes da aviação e desenvolvido para outros domínios técnicos (Ibidem, pp. 55-56).

Cabe também mencionar o *Multinational Customized English* (MCE), a linguagem controlada utilizada pela Xerox Corporation e aplicada ao sistema de tradução automática Systran. Graças à implementação desta linguagem controlada no sistema, os resultados mostraram uma redução das ambiguidades no *input* e, ao mesmo momento, uma melhoria do *output*. Além disso, os textos redigidos são de fácil compreensão e permitem traduções rápidas para línguas diferentes (Elliston, 1979 *apud* Banjar, 2001: 59).

As linguagens controladas, hoje em dia, podem ser utilizadas também na perspectiva da aprendizagem de línguas estrangeiras e/ou para redigir determinados tipos de textos com vista a uma melhor compreensão por parte de não nativos. Na secção seguinte, são classificados os vários tipos de linguagens controladas e são apresentadas as características que as distinguem.

3.2 TIPOLOGIA DE LINGUAGENS CONTROLADAS

Em termos gerais, as linguagens controladas podem ser subdivididas conforme o tipo de problemas que se pretende resolver (Schwitter, 2002 *apud* Kuhn, 2014). No essencial, assumem-se aqui as distinções descritas em Kuhn (2014).

Huijsen (1998), entre outros, distingue entre *Human oriented controlled language* (HOCL) e *Machine-oriented controlled language* (MOCL). As primeiras visam melhorar a legibilidade e a compreensão de um determinado texto, enquanto as segundas têm como objetivo melhorar o processo de tradução. Contudo, ainda segundo o mesmo autor, a distinção entre HOCL e MOCL nem sempre se pode estabelecer de maneira clara, pois a simplificação de um texto pode funcionar para ambas as finalidades.

Outra distinção baseia-se no tipo de abordagem – em concreto, abordagem “naturalista” e abordagem “formalista” (Marrafa *et al.*, 2012; Clark *et al.*, 2009). A abordagem “naturalista” interpreta as LC como conjuntos de restrições sobre léxico e estruturas já existentes numa dada língua natural e estabelece quais unidades lexicais e estruturais não devem ser utilizadas. A abordagem “formalista” considera as linguagens controladas como um conjunto de regras e de vocabulário para a formação de expressões numa dada língua natural, as quais determinam o léxico e as estruturas permitidas (Marrafa *et al.*, 2012).

Reuther (2003) considera outra distinção, baseada nos conceitos de legibilidade e traduzibilidade. De acordo com a autora, a abordagem por legibilidade visa simplificar a linguagem para o texto ser lido sem nenhum tipo de ambiguidade por leitores quer nativos quer não (como no caso do *Basic English*). No outro tipo de abordagem, o principal objetivo é o melhoramento dos resultados de tradução, em particular no caso da tradução automática.

Em relação às regras para a linguagem controlada, classificam-se geralmente em dois tipos: podem ser proscritivas ou prescritivas, ou também uma combinação das duas (Nyberg, Mitamura, Huijsen, 2003 *apud* Kuhn, 2014). As regras proscritivas descrevem o que não é permitido, enquanto as regras prescritivas descrevem o que é. Também pode haver uma combinação entre os dois tipos de regras, a qual pode determinar um tipo de linguagem mais específica ou restrita a um determinado par linguístico.

Em resumo, as linguagens controladas podem ser classificadas conforme os seus objetivos, mas também conforme a maneira como são estruturadas e conforme as regras que são criadas.

3.3 VIRTUALIDADES E LIMITES DAS LINGUAGENS CONTROLADAS

A alteração dos elementos do texto de partida com a utilização de uma linguagem controlada leva a muitas vantagens sob vários aspetos.

Para as HOCL uma das vantagens maiores é a transparência dos conteúdos, para que as informações sejam simples e claras para o maior número possível de pessoas (em particular para os utilizadores de manuais ou textos de tipo técnico). Além disso, também podem ser úteis para pessoas não nativas da língua ou com dificuldades cognitivas de qualquer tipo. Contudo, neste tipo de trabalho não cabe o tratamento deste tipo de LC, pois as que são aqui relevantes são as MOCL, dado o âmbito deste trabalho ser o da tradução automática.

A vantagem principal das linguagens controladas de tipo MOCL é, obviamente, a obtenção de melhores resultados de tradução em comparação com textos escritos em linguagem natural. Adicionalmente, uma vez que em tais circunstâncias será precisa apenas a revisão do texto, ou pós-edição, o controlo do texto de *input* representará ainda uma poupança em termos de tempo e provavelmente também em termos financeiros.

No entanto, dificilmente existirão linguagens controladas que forneçam um conjunto de regras exaustivo que possa lidar com todos os fenómenos de uma determinada língua. A dificuldade de captação de todas as especificidades da língua em contraste com a outra língua envolvida torna difícil a exaustividade das regras para a linguagem controlada. O sistema de tradução automática em questão tem de ser estudado de maneira aprofundada e os diversos problemas de tradução derivantes (ou não) de tais especificidades devem ser tidos em consideração, de maneira a encontrar regras cuja aplicação permita evitá-los. Um exemplo prático, como se vê mais à frente, são as preposições que introduzem locativos em italiano, pois por serem diferentes do português causam consideráveis problemas à tradução automática e muitos desafios à criação de regras para a linguagem controlada.

Além disso, as regras para uma determinada linguagem controlada são, com alguma frequência, estabelecidas com base num determinado sistema de tradução automática e muitas vezes para um determinado par linguístico, o que implica que nem sempre podem ser aplicadas a outros sistemas e/ou outros pares linguísticos.

As linguagens controladas que são inseridas nos sistemas de tradução automática, no entanto, têm características diferentes relativamente às que se acabam de referir.

Este tipo de controlo automático é, naturalmente, adaptado às características do sistema de TA que integra, respeitando ainda opções dos construtores em termos de tradução e da própria escrita. Uma das características que muitas destas linguagens controladas têm em comum é a redução do vocabulário e/ou a integração de dicionários terminológicos no sistema.

No caso do já mencionado *Caterpillar Fundamental English*, por exemplo, o vocabulário foi limitado a 850 palavras. Sucessivamente, quando se tornou *Caterpillar Technical English*, foi integrado um dicionário terminológico e foram escolhidos 70.000 termos (gerais e técnicos) entre um milhão para serem utilizados. Desta maneira, a redução do vocabulário permitiu ao sistema uma maior precisão na escolha dos termos. Além do mais, graças à sua eficiência a nível terminológico, o *CTE* permitiu aos autores reutilizar os mesmos termos para os vários produtos da empresa e, por conseguinte, aumentou a produtividade.

Assim sendo, as linguagens controladas implementadas no sistema, em geral, têm a vantagem de permitir a obtenção de textos mais coerentes e mais consistentes, seja a nível de estilo, de conteúdo e de terminologia. Por outro lado, podem causar algumas dificuldades por parte dos autores, pois algumas regras para a escrita podem entrar em conflito com as regras para o melhoramento do desempenho do sistema (Cremers, 2011, *apud* Zamagni, 2015: 63).

3.4 INTRODUÇÃO À LINGUAGEM CONTROLADA PARA TRADUÇÃO AUTOMÁTICA DO PAR ITALIANO-PORTUGUÊS

As linguagens controladas, como referido acima, são constituídas por restrições, apresentadas sob o formato de regras, a serem aplicadas a textos em linguagem natural. Citando Arnold *et al.* (1994: 26),

“[e]xactly what constitutes good input will vary a little from system to system. However, it is easy to identify some simple writing rules and strategies that can improve the performance of almost any general-purpose MT system.”

Entre as regras mencionadas pelo mesmo autor, no presente trabalho são tidas em consideração principalmente as seguintes¹⁰:

- “Keep sentences short.”
- “Avoid (so far as possible) words which have several meanings.”

Conforme exposto em tais sugestões, é preciso usar frases curtas, sendo que, geralmente, os sistemas de tradução automática têm dificuldade em analisar frases longas e logo em traduzi-las apropriadamente. Contudo, esta regra geral pode entrar em conflito com alguma regra específica, dependendo da complexidade da frase envolvida. Se uma frase longa for menos complexa e for, por conseguinte, de maior eficiência para o desempenho do sistema do que uma frase curta semanticamente equivalente, haverá que optar pela primeira.¹¹

As palavras que têm uma forma para dois ou mais sentidos devem ser evitadas porque podem causar ambiguidade lexical (como explicado na secção 2.4). Por isso, é preciso pensar bem nas possíveis interpretações que pode ter uma palavra ou expressão e, eventualmente, escolher um sinónimo que não seja ambíguo.

O utilizador precisará considerar tanto as regras gerais como as específicas, de modo que as restrições mais relevantes tenham precedência sobre as outras.

Como já referido anteriormente, o sistema de tradução automática utilizado para a elaboração e teste das regras para a linguagem controlada é a versão *online* do Systran, SystraNet.

A abordagem utilizada para este tipo de LC visa, além do mais, a consecução de *inputs* não ambíguos, tendo em conta quer a ambiguidade lexical quer a estrutural. Segundo a classificação de Huijsen, entre outros, é uma linguagem controlada de tipo MOCL (*Machine-Oriented Controlled Language*), pois tem como objetivo o incremento da qualidade dos resultados da tradução automática, que, no caso vertente, envolve o par italiano > português. Por fim, as regras são uma combinação entre proscritivas e prescritivas, pois através das proscritivas define-se o que não permite um bom desempenho do sistema e através das prescritivas define-se o que melhora o seu desempenho.

¹⁰ As regras mencionadas pelo autor que conflituam com o princípio adotado no presente trabalho não são aqui citadas.

¹¹ A complexidade da frase não está necessariamente relacionada com a sua extensão.

Em seguida, são expostos alguns fragmentos de linguagem controlada aplicados para o par italiano-português, que exemplificam a aplicação de algumas restrições gerais.

Quando o *input* envolve frases com constituintes nulos e pode criar-se uma ambiguidade em termos de concordância, é preciso especificá-los, embora ambas as línguas italiano e português não precisem obrigatoriamente do sujeito antes do verbo. Tome-se em consideração esta frase não controlada:

(1) Maria ha detto a Giovanni che \emptyset è stanca. > *Marie disse Giovanni que é cansado.

O resultado da tradução fica uma frase agramatical por não ter concordância de género entre o sujeito, que é feminino (“Maria”), e o adjetivo, que é masculino (“cansado”). Esta falta de concordância é devida ao facto de o verbo *essere* não ter o sujeito especificado, pelo que o sistema não associa o sujeito de género feminino diretamente ao adjetivo. Como já mencionado anteriormente (secção 2.4), trata-se de ambiguidade do sistema, sendo que o ser humano consegue perceber que o sujeito do verbo *essere* é *Maria*, mas o sistema não tem esta informação.

Assim sendo, em casos parecidos com (1), é preciso explicitar o pronome antes do verbo, para assegurar-se que o sistema reconheça esta concordância:

(1) a. Maria ha detto a Giovanni che lei è stanca. > *Marie disse Giovanni que é cansada.¹²

Como se pode notar, apesar dos erros sintáticos da frase, o controlo da linguagem permite que o sistema identifique o género do adjetivo (“cansada”).

Quando no *input* occorem itens lexicais ambíguos, é preciso substituí-los por outros não ambíguos. Tome-se o exemplo da ambiguidade de *pesca*, a que estão associados dois conceitos: pêssego e pesca:

(2) Mario va a pesca tutti i giorni. > Mario vai à pesca todos os dias.

(3) Mario mangia molte pesche. > *Mario come muitas pescas.

Neste caso, não há desambiguação no sistema pelo facto de a palavra *pesca* ser traduzida em ambos os casos para *pesca*, o que torna (3) agramatical em português. A desambiguação deve ser efetuada através da substituição por equivalentes, tendo em

¹² Apesar de a tradução ser correta do ponto de vista da concordância sujeito-verbo, ficam problemas devidos ao uso de *ser* e *estar* e à falta da preposição *a* depois do verbo *dizer*.

conta qual dos dois conceitos associados a *pesca* é escolhido pelo sistema. No caso de *pesca*, é reconhecida apenas a palavra que designa o desporto, pelo que é preciso substituir a palavra por um sinónimo, hipónimo ou hiperónimo de *pêssego*:

(3) a. Mario mangia molta frutta. > Mario come muitos frutos¹³.

(3) b. Mario mangia molte pesche noci. > Mario come muitas nectarinas.

Como a palavra *pesca* (“pêssego”) não tem sinónimos, é substituída por um hiperónimo na frase (3) a., e por um hipónimo na frase (3) b. Contudo, neste trabalho não cabe o estabelecimento de qual das duas opções seja melhor para o controlo da linguagem, pois em nenhum dos dois casos o resultado de tradução é a palavra desejada. A escolha entre hipónimo e hiperónimo depende do contexto em que a frase se insceva.

Por último, cabe tomar em consideração um exemplo que demonstre a importância da utilização de frases curtas:

(4) La Commissione e le aziende informatiche sono tuttavia consapevoli del fatto che la diffusione dell'illecito incitamento all'odio online si ripercuote negativamente non solo sui gruppi o sui singoli che vengono presi di mira, ma anche su coloro che nelle nostre società aperte si esprimono a favore della libertà, della tolleranza e della non discriminazione, e ha un effetto inibitore sul discorso democratico sulle piattaforme online.¹⁴

O resultado da tradução automática é o seguinte:

(4) a. A Comissão e as empresas informáticas são contudo conscientes do facto que a divulgação do ilícito incentivo ao ódio em linha reflete-se negativamente não **somente sobre** os grupos ou **sobre** os indivíduos que são **tomados de mira, mas mesmo** sobre os que nas nossas sociedades abertas exprimem-se em prol da liberdade, a tolerância e o passo discriminação, e **tidos** um efeito **inibiteur sobre** o discurso democrático sobre as plataformas em linha.*¹⁵

¹³ A tradução correta é *fruta*, sendo que *fruto* é o termo da biologia e só se usa em algumas expressões (“frutos secos”, “frutos silvestres”, etc.). Contudo, o resultado é pouco relevante para o caso.

¹⁴ Exemplo tirado do artigo *Odio online: Commissione europea presenta un codice di condotta*, na revista jurídica italiana “Altalex”, <http://www.altalex.com/documents/news/2016/05/31/odio-online-commissione-europea-presenta-con-codice-di-condotta>.

¹⁵ Em negrito são evidenciados os elementos de agramaticalidade/inadequação.

Para o controlo da linguagem deste texto é preciso ter em consideração vários aspetos, de que se destacam:

1. A pontuação.
2. As expressões idiomáticas e as colocações da língua de *input*.
3. A estrutura da frase de *input* e da de *output*.

Para controlar a pontuação, é preciso usar sinais de pontuação mais fortes¹⁶ para formar frases mais simples e menos compridas. Desta maneira, o sistema conseguirá melhor desempenho, sendo os resultados de melhor qualidade.

Para o controlo das expressões idiomáticas e das colocações, é preciso substituí-las por expressões de sentido literal. Neste caso, a expressão *sono presi di mira* pode ser substituída por *sono i bersagli* (“são os alvos”).

Tendo em conta que as duas línguas têm especificidades estruturais, é preciso adaptar algumas estruturas em função das da língua de *output*.

Assim, à luz destas considerações, será efetuado o controlo da linguagem:

(4) b. La commissione e le aziende informatiche, tuttavia, sono consapevoli di **un fatto**: \emptyset la diffusione dell'illecito incitamento all'odio online si ripercuote negativamente **in** gruppi o singoli, **che sono i bersagli, e nelle persone che si esprimono**, nelle nostre società aperte, a favore di **valori come** \emptyset libertà, \emptyset tolleranza e \emptyset **no-discriminazione**. **Questo** ha un effetto **di inibizione nel** discorso democratico sulle piattaforme online.¹⁷

O resultado da tradução automática é o texto seguinte:

(4) c. A comissão e as empresas informáticas, contudo, são¹⁸ conscientes de um facto: a divulgação do ilícito incentivo ao ódio em linha reflete-se negativamente em grupos ou indivíduos que são os alvos e nas pessoas que se exprimem, nas nossas sociedades abertas, em prol de valores como liberdade, tolerância e não discriminação. Isto tem um efeito de inibição no discurso democrático sobre as plataformas em linha.

Como se pode observar em (4) c., o resultado da tradução já não é agramatical. Assim, as frases ficam mais curtas pela introdução de pontuação (ponto, vírgulas e dois pontos) e do demonstrativo anafórico *questo*. A sintaxe é ligeiramente alterada para permitir a

¹⁶ Sinais de pontuação mais fortes, neste caso, são ponto, ponto e vírgula, dois pontos e acrescentamento de vírgulas.

¹⁷ Em vermelho são evidenciadas as modificações da frase em linguagem controlada.

¹⁸ O correto é “estão”. Não se chegou a uma forma eficaz de controlo.

adequação do enclítico sem o complementador *que* (\emptyset “reflete-se”) e a ocorrência do proclítico (“que se exprimem”). A locução conjuncional copulativa afirmativa *non solo...ma anche* é substituída apenas pela conjunção copulativa *e*¹⁹, para simplificar a frase.

Por causa da diferença de uso das preposições nas duas línguas, *su* é substituído por *in*, conforme a regência do verbo. Além da substituição da expressão *sono presi di mira* por *sono i bersagli*, também se substitui o pronome *coloro* por *le persone*, por ser uma expressão mais facilmente reconhecível pelo sistema. Além disso, *inibitore* é substituído pelo substantivo *inibizione* porque a tradução do adjetivo é errada.

Contudo, através da linguagem controlada, muitas vezes, é inevitável a degradação da língua de *input*, posto que deve ser adaptada à língua de *output*. No entanto, o que mais interessa nas linguagens controladas orientadas para a tradução automática é que o resultado de tradução seja gramatical, pelo que a aceitabilidade da língua de *input* não tem nenhuma importância.

Nos capítulos a seguir são tratadas as problemáticas centrais deste trabalho, na base de uma análise contrastiva da expressão do modo, da modalidade e do aspecto, com suporte empírico em textos do italiano (maioritariamente autênticos) e respectivas traduções para português através do SystraNet. São igualmente inventariados alguns dos fenómenos encontrados nestas análises e tratados separadamente, num capítulo à parte (capítulo 7).

Há igualmente que mencionar o problema da tradução dos nomes próprios de pessoa, os quais são quase sempre traduzidos pelo sistema. Dado que este tipo de controlo não é possível, os resultados de tradução dos nomes próprios pelo sistema são ignorados e assumidos como aceitáveis. A única maneira de efetuar um controlo deste tipo de problemas seria aceder à opção “My Dictionary” e criar uma conta personalizada no sistema; contudo, o tratamento deste tipo de controlo não cabe nos objetivos deste trabalho.

¹⁹ Aplicação da regra 10 de Annalisa Zamagni (2015).

4. MODO

Em termos gerais, o modo é a categoria gramatical dos verbos que transmite a atitude do falante, relacionando-se com a modalidade em aspetos vários. O modo imperativo relaciona-se com a modalidade deôntica, na medida em que expressa – na maioria dos casos – uma ordem. Em contrapartida, os outros modos – nomeadamente, o indicativo e o conjuntivo – relacionam-se com a modalidade epistémica (Squartini, 2010: 111).

Quer em italiano quer em português, consideramos basicamente os modos acima referidos: imperativo, indicativo e conjuntivo. Contudo, o condicional também pode ocorrer como modo, embora ocorra fundamentalmente como um tempo (cf. entre outros, Oliveira, 2003).

O mais relevante na variação de modo respeita naturalmente à sua diferente distribuição nas duas línguas. Em relação ao imperativo, registam-se bastantes diferenças, o que provoca diversos tipos de problemas na tradução automática. Estas diferenças são analisadas detalhadamente com o fim de estabelecer um conjunto de regras com a maior cobertura possível para a linguagem controlada.

Neste capítulo trata-se do modo e das suas especificidades em italiano e em português, evidenciando todas as falhas do sistema durante a tradução.

Em primeiro lugar, é tratado o modo imperativo, ou melhor, as chamadas frases imperativas, salientando todas as diferenças que causam problemas ao sistema e para as quais teremos de encontrar as soluções mais adequadas.

Em seguida, são tratados os casos em que o modo conjuntivo ocorre em orações independentes/frases simples, embora seja bastante rara a sua ocorrência na língua italiana neste tipo de frases.

Por último, é analisada a distribuição do indicativo e do conjuntivo nas frases completivas e adverbiais, pois os dois modos ocorrem em contextos diferentes em italiano e em português e, portanto, é preciso proceder ao pertinente controlo da linguagem.

4.1 FRASES IMPERATIVAS

O modo imperativo tem uma distribuição diferente nas duas línguas, tanto nas imperativas afirmativas como nas imperativas negativas.

Em italiano, o verbo da frase imperativa afirmativa na segunda pessoa do singular tem a forma igual à da segunda pessoa do singular do presente do indicativo, no caso dos verbos de segunda e terceira conjugação. No caso dos verbos de primeira conjugação, a forma usada é a da terceira pessoa do singular do indicativo:

Infinitivo do verbo	Segunda pessoa singular do imperativo	Formas iguais do presente do indicativo
<i>mangiare</i> (1ª conjugação) 'comer'	mangia!	(lui) mangia
<i>temere</i> (2ª conjugação) 'temer'	temi!	(tu) temi
<i>sentire</i> (3ª conjugação) 'sentir', 'ouvir'	senti!	(tu) senti

Tabela 1. Analogia entre formas do imperativo e do indicativo.

A forma de segunda pessoa do plural do imperativo dos verbos regulares é a da segunda pessoa do plural no presente do indicativo:

mangiate!	(voi) mangiate
temete!	(voi) temete
sentite!	(voi) sentite

Tabela 2. Analogia entre formas do imperativo e do indicativo.

Em vez da segunda pessoa do plural do indicativo, com alguns verbos irregulares, especificamente verbos auxiliares, faz-se uso da correspondente forma de conjuntivo:

Infinitivo do verbo	Segunda pessoa plural do imperativo	Formas iguais do presente do conjuntivo
avere 'ter'	abbiate!	che voi abbiate
essere 'ser'	siate!	che voi siate

Tabela 3. Analogia entre formas do imperativo e do indicativo.

Em português, a frase imperativa afirmativa tem o verbo no imperativo da segunda pessoa do singular, que tem a mesma forma da terceira pessoa do presente do indicativo.

A segunda pessoa do plural é muito pouco utilizada na língua contemporânea, pelo que não será tida em consideração. Para os pronomes *você* ou *vocês*, na frase imperativa utiliza-se o verbo conjugado na correspondente terceira pessoa do presente do conjuntivo.

A frase imperativa negativa tem uma estrutura muito simples na língua italiana: para a segunda pessoa do singular basta usar a negação *e*, a seguir, o verbo no infinitivo (*non* + infinitivo); para a segunda pessoa do plural a forma do verbo é a do presente do indicativo (ex. *non mangiate!*), com exceção dos verbos irregulares que seguem a forma do conjuntivo (ex. *non abbiate!*), como já referido.

Na frase imperativa negativa do português, o modo imperativo é substituído pelo conjuntivo.

Tanto as frases imperativas afirmativas como as negativas podem exprimir uma ordem, um comando, um conselho, uma súplica ou simplesmente um desejo. Neste último caso, em ambas as línguas é aceite o verbo no conjuntivo presente, o qual adquire um valor de tipo exortativo.

Na frase imperativa afirmativa, o SystraNet não reconhece a segunda pessoa do singular do imperativo e identifica-a como a segunda pessoa do singular do presente do indicativo. Em frases que expressam desejo, como no caso abaixo, o controlo da linguagem pode ser efetuado com o conjuntivo:

(1) Dormi bene! > *Dormes bem!

LC:

(1) a. Che tu dorma bene! > Que durmas bem!

(2) Fai buon viaggio! > *Fazes bom viagem!

LC:

(2) a. Che tu faccia buon viaggio! > *Que tu fazes bom viagem!

(2) a. não é a escolha mais adequada para o controlo da linguagem porque o verbo no conjuntivo é traduzido pelo verbo correspondente no indicativo. Isto acontece porque o sistema não reconhece as formas de conjuntivo de alguns verbos, como é o caso de *fare*:

che io faccia	que faço
che tu faccia	que fazes
che lui faccia	que faz
che noi facciamo	que fazemos
che voi facciate	*que vocês fachadas
che loro facciano	que façam

Tabela 4. Presente do conjuntivo do verbo *fare* e resultados de tradução do SystraNet.

A única forma verbal devidamente reconhecida pelo sistema é a da terceira pessoa do plural (“que façam”), a segunda pessoa do plural até é traduzida pelo substantivo *fachadas* por ambiguidade lexical (*facciate* plural de *facciata*). Todas as outras formas são traduzidas no indicativo.

Além disso, há uma falta de concordância entre o adjetivo e o substantivo (*viagem* é feminino e *bom* é masculino) que tem de ser tratada.

Por conseguinte, embora em (1) a. o controlo seja bem sucedido, não é possível estabelecer a regra que implica a substituição do imperativo pelo conjuntivo precedido por *che*, sendo que não resulta eficaz no caso de (2) a.

Assim sendo, a solução mais adequada consiste em substituir a forma imperativa italiana pela terceira pessoa do presente do indicativo, especificando também o sujeito. Desta forma, obtém-se o imperativo, segunda pessoa, em português:

(1) Dormi bene!

LC:

(1) b. Lui dorme bene! > Dorme bem!

(2) Fai buon viaggio!

LC:

(2) b. Lui fa un buon viaggio!²⁰ > Faz uma boa viagem!

Às frases imperativas que expressam ordem ou conselho, consequentemente, aplica-se o mesmo controlo:

(3) Prendi un’aspirina! > *Tomas aspirina!

²⁰ Controlo através de uma frase em que o sintagma nominal objeto integra o determinante *un*.

LC:

(3) a. Lui prende una aspirina! > Toma aspirina!

(4) Vieni qui! > *Vens aqui!

LC:

(4) a. Lui viene qui! > Vem aqui!

(5) Andate a casa! > *Aleia à casa!

LC:

(5) a. Loro vanno a casa! > Vão à casa!

LC:

(5) b. Loro vanno la casa!²¹ > Vão a casa!

Na frase (5), o sistema nem sequer reconhece o verbo na segunda pessoa do plural do presente do indicativo, razão pela qual a linguagem é controlada substituindo-se a segunda pela terceira pessoa do plural (frase (5) a.).

(6) Portami un caffè! > *Esticado mim um café!

‘traz-me um café!’

Quando o verbo tem um pronome clítico associado, na maioria dos casos o sistema não consegue reconhecer o verbo. Provavelmente este fenómeno é devido ao facto de em italiano não haver separação entre o verbo e o pronome, enquanto em português verbo e pronome são separados por um hífen (*portami* > *traz-me*).

Evitando estes pronomes ou “desamalgamando-os” do verbo, o sistema dá resultados diferentes:

(6) a. Porta a me un caffè! > *Leva à mim um café!

(6) b. Porta un caffè! > *Esticado um café!

No primeiro caso, o resultado de tradução da frase (6) a. é interpretável como equivalente à frase de partida, mas a não escolha do clítico no caso apropriado, bem como o facto de ser precedido de determinante (*a*), faz com que a expressão seja

²¹ A expressão aceitável é “ir a casa” e não “ir à casa”, pelo que o controlo da linguagem será efetuado com o determinante definido no sintagma nominal, não precedido de preposição, *la casa*, para que na tradução ocorra só a preposição (que tem a mesma forma que o artigo em português). Este tipo de controlo será aprofundado no capítulo 7.

agramatical. Todavia, o sistema reconhece o verbo *portare*, mas traduz por *levar* em vez de *trazer*, uma vez que *portare* é um caso de ambiguidade lexical. Não havendo possibilidade de desfazer essa ambiguidade, a única forma de evitar o problema é encontrar uma paráfrase que não inclua *portare*.

Ao tirar o pronome, no caso da frase (6) b., o sistema continua sem reconhecer o verbo *portare*, pelo que a solução melhor será, também aqui, realizar o sujeito:

LC:

(6) c. Lui porta un caffè! > Leva um café!

Em relação à frase imperativa negativa da segunda pessoa do singular, o que se afigura como mais natural é controlar a linguagem substituindo o verbo no infinitivo pelo conjuntivo presente, com vista a reproduzir as estruturas imperativas negativas em português:

(7) Non lasciare la luce accesa! > *Não deixar a luz acendida!²²

LC:

(7) a. Non lasci la luce accesa! > *Não lasci a luz acendida!

O sistema não só tem problemas em reconhecer o conjuntivo, mas também não reconhece algumas formas de indicativo do verbo *lasciare* (“deixar”):

io lascio	deixo
tu lasci	*você lasci
lui lascia	deixa
noi lasciamo	*nós lasciamo
voi lasciate	*deixa
loro lasciano	deixam

Tabela 5. Presente do indicativo do verbo *lasciare* e resultados de tradução do SystraNet.

Como se pode observar, só algumas formas são reconhecidas pelo sistema: a segunda pessoa do plural é traduzida pela terceira pessoa do singular e a primeira e segunda do

²² A tradução (à parte participio regular) seria correta se o destinatário tivesse referência arbitrária, mas neste caso referimo-nos à segunda pessoa do singular.

Quanto ao participio, das duas formas, *acendida* ou *acesa*, o sistema escolhe a primeira por ser regular, mas a que se deve usar neste caso é a irregular, por se tratar do participio passivo. Havendo em italiano só uma forma de participio, não há aqui possibilidade de controlo.

singular são totalmente ignoradas pelo sistema. Acontecendo que para o sistema estas formas não existem, não se pode encontrar uma solução para que sejam traduzidas.

A única maneira de resolver o problema que se coloca na tradução da frase imperativa negativa será encontrar uma frase semântica e pragmaticamente equivalente ou quasi-equivalente que não tenha a construção infinitiva nem o verbo no conjuntivo, por exemplo:

LC:

(7) b. Non devi lasciare la luce accesa. > *Não debes deixar a luz acendida.

Veja-se outro exemplo:

(8) Non mi portare a casa! > *Não me levar à casa!

LC:

(8) a. Non mi porti a casa! > *Você mim portas não à casa!

No caso da frase (8) a., o sistema altera completamente a ordem da frase, colocando a negação depois do verbo e o pronome *voce* no início da frase. O verbo *portare* é traduzido por *portar* em vez de *levar* e conjugado no indicativo presente da segunda pessoa do singular. Além disso, o objeto pronominal, *mi*, é traduzido por *mim*, em vez de por *me*.

Contudo, também neste caso de frase imperativa negativa não há hipótese de controlar a linguagem e a única solução, uma vez mais, é substituir a imperativa por uma expressão equivalente:

LC:

(8) b. La mia richiesta è di non portarmi a casa. > O meu pedido é não me levar à casa.

LC:

(8) c. La mia richiesta è di non portarmi la casa.²³ > O meu pedido é não me levar a casa.

(9) Non legga quell'articolo! > *Não lê este artigo!²⁴

'Não leia aquele artigo!'

²³ *a casa*, fica *la casa* para permitir uma correta tradução, *a casa* (ver também nota 21).

²⁴ O sistema não reconhece a correta correspondência dos adjetivos demonstrativos: *quello* corresponde a *aquilo* e *questo* corresponde a *este*. Em italiano não existe a forma de demonstrativo que em português é expressa por *esse*, portanto esta diferença causa problemas à tradução automática.

A estrutura da frase imperativa negativa para a terceira pessoa do singular é a mesma em português e em italiano. Mesmo assim, o sistema continua a apresentar problemas na tradução de alguns verbos no conjuntivo presente:

che io legga	que leio
che tu legga	*que você lises
che lui legga	*que lê
che noi leggiamo	*que lemos
che voi leggiate	*que leia
che loro leggano	*que leem

Tabela 6. Presente do conjuntivo do verbo *leggere* e resultados de tradução do SystraNet.

Como se pode observar, o sistema não reconhece todas as formas do conjuntivo presente do verbo *leggere* (“ler”). De facto, as primeiras e as terceiras pessoas (do singular e do plural) são traduzidas pelas formas correspondentes no presente do indicativo, a segunda do singular é traduzida pela forma do conjuntivo presente francês (provavelmente pela utilização como língua ponte) e a segunda do plural é traduzida pela terceira do singular.

Neste caso, sendo que se trata de uma recomendação, poderá expressar-se o mesmo sentido da frase usando uma expressão equivalente, como por exemplo:

LC:

(9) c. Non le raccomando di leggere quell’articolo. > *Não o recomendo deste ligeiros²⁵ artigo.

O pronome objeto indireto *le* é traduzido pelo pronome objeto direto. Assim, para que a tradução resulte correta, há que substituí-lo pelo pronome objeto indireto de terceira pessoa:

LC:

(9) d. Non gli raccomando la lettura di quell’articolo. > Não lhe recomendo a leitura deste artigo.

²⁵ Há aqui uma ambiguidade lexical entre *leggere* (verbo “ler”) e *leggere* (adjetivo feminino plural “ligeiras”).

4.2 CONJUNTIVO EM FRASES SIMPLES

Como evidenciado na secção precedente, o modo conjuntivo pode ocorrer, tanto em italiano como em português, em frases simples com valor imperativo. Em italiano é mais raro do que em português encontrar o conjuntivo em outros tipos de frase simples. Um dos casos em que o conjuntivo pode ocorrer em italiano é em frases interrogativas globais com valor dubitativo:

“[i]n independent non-declarative clauses the Subjunctive can also acquire an epistemic meaning (dubitative conjectures), occurring in interrogative clauses with mandatory complementizer *che*” (Squartini, 2010: 247).

Uma forma de expressar este tipo de dúvida em português é através do futuro do indicativo. Este uso do futuro do indicativo verifica-se também na língua italiana e exprime, tanto em italiano como em português, uma dúvida do falante em relação a um determinado tema. Sendo assim, o futuro pode substituir o conjuntivo neste contexto e, por conseguinte, servir para controlar a linguagem em frases como as que se seguem:

Presente do conjuntivo:

(10) *Che sia una bugia?* > **Que seja uma mentira?*

Futuro simples:

(10) a. *Sarà una bugia* > *Será uma mentira?*

Em português, diversamente do que acontece em italiano, o modo conjuntivo ocorre nas frases imperativas negativas e também nas terceiras pessoas das imperativas afirmativas. Contrariamente ao italiano, em português o modo conjuntivo pode ocorrer também em vários tipos de frase simples. Entre outras, cabe mencionar a sua ocorrência depois de expressões de desejo (por exemplo, *oxalá* ou *quem me dera*) ou em frases dubitativas introduzidas por *talvez*:

(11) Talvez a Giulia esteja no escritório.

A frase correspondente em italiano, contrariamente ao português, tem o verbo no indicativo:

(11) a. Forse Giulia è in ufficio.

Assim, para o controlo da linguagem, pode tomar-se em consideração a opção de conjugar o verbo italiano no conjuntivo, para que a tradução para português resulte também no conjuntivo:

LC:

(11) b. Forse Giulia sia in ufficio. > *Talvez Julie esteja escritório.

LC:

(11) b. Forse Giulia sia nell'ufficio.²⁶ > Talvez Julie esteja no escritório.

Vejam-se outros exemplos:

(12) Forse tu non conosci mia sorella.

LC:

(12) a. Forse tu non conosca mia sorella. > * Talvez não conheces a minha irmã

Embora em (12) a. o verbo esteja no conjuntivo presente, o sistema não consegue traduzi-lo de maneira correta.

Vejam-se também frases com o verbo no passado:

(13) Forse Giovanni e Luca non hanno cenato.

LC:

(13) a. Forse Giovanni e Luca non abbiano cenato. > *Talvez Giovanni e Luca não jantaram.

Como demonstram estes exemplos, o sistema não consegue traduzir os verbos no presente e no passado do conjuntivo. Nos casos de (12) a. e (13) a., tal acontece porque o sistema não reconhece estas formas verbais como formas de conjuntivo neste tipo de construções sintáticas.

Dado que não se pode prever se o sistema traduzirá corretamente os verbos no conjuntivo e tendo em conta que tal fenómeno se verificou em diversas situações, a solução melhor é deixar o verbo no indicativo e colocar o advérbio a seguir, ainda que esta não seja a ordem preferencial:

²⁶ Controlo da linguagem com a *preposizione articolata* (“nell'ufficio”) para que o sistema traduza a preposição em português.

LC:

(11) d. Giulia è nell'ufficio, forse. > Julie está no escritório, talvez.

LC:

(12) b. Tu non conosci mia sorella, forse. > Tu não conheces a minha irmã, talvez.

LC:

(13) b. Giovanni e Luca non hanno cenato, forse. > Giovanni e Luca não jantaram, talvez.

4.3 FRASES COMPLETIVAS

A subordinação completiva caracteriza-se pelo facto de a frase subordinada constituir um argumento da frase superior. Em português, as frases completivas podem ser finitas (introduzidas por complementador (*que*) e com o verbo no conjuntivo ou no indicativo) ou não finitas (sem complementador e com o verbo no infinitivo, flexionado ou não flexionado). Em italiano estes dois tipos de completivas chamam-se comumente *esplicita* (introduzidas por complementador (*che*) e com o verbo no conjuntivo ou no indicativo) e *implicita* (introduzidas por *di* e com o verbo no infinitivo). Em italiano a *implicita* é aceite só no caso de o sujeito da subordinada ser co-referente com o da principal. Pelo contrário, em português usa-se o infinitivo pessoal (o qual não tem correspondência em italiano), pelo que os sujeitos da principal e da subordinada podem ter referência disjunta.

FRASES COMPLETIVAS FINITAS.

A seleção do modo nas frases completivas finitas, tanto em italiano como em português, depende de diferentes fatores, em particular da natureza semântica do verbo de que dependem, sendo que verbos conceptualmente idênticos em italiano e em português nem sempre selecionam o mesmo modo na completiva, o que traz problemas à tradução.

O uso do conjuntivo em italiano, muitas vezes, causa problemas ao sistema. De facto, nas frases completivas, o sistema dificilmente reconhece o conjuntivo, seja presente ou passado. O controlo da linguagem, deve, por conseguinte, ter em conta não apenas as diferenças linguísticas entre italiano e português, mas também vários aspetos respeitantes ao deficiente desempenho do sistema.

Seguem alguns exemplos de seleção do modo do SystraNet:

(14) Pensavo che fosse troppo presto. > Pensava que era demasiado cedo.

(15) Penso che sia troppo presto. > Penso que seja demasiado cedo.

(16) Credo che sia troppo presto. > Acredito que seja demasiado cedo.

Em todos estes exemplos (14-16), contrariamente ao italiano, que admite o conjuntivo²⁷, a seleção do modo em português depende do grau de “crença” por parte do locutor (Marques, 2010: 151).²⁸ Assim sendo, nestes casos, quer o indicativo (14) quer o conjuntivo (16-17) são aceites na língua portuguesa e, em consequência, não é preciso controlar a linguagem.

(17) Trovo che Maria stia meglio. > ? Encontro que Marie esteja melhor²⁹.

No exemplo (17), o fator de fraca aceitabilidade da tradução é o verbo *trovare*, cuja tradução, neste contexto, deve ser *achar*. Dado que o sistema não identifica este verbo como correspondente de *achar*, traduz o verbo *trovare* por *encontrar*, que não se adequa ao contexto, salvo num registo muito popular. Um verbo que pode expressar o mesmo sentido é o verbo *pensar*, pelo que deve usar-se *pensare* em vez de *trovare*:

(17) Trovo che Maria stia meglio.

LC:

(17) a. Penso che Maria stia meglio. > Penso que Marie esteja melhor.

LC:

(17) b. Penso che Maria sta meglio. > Penso que Marie está melhor.

Como se pode notar em (17) a. e (17) b., o sistema reconhece os dois modos do verbo *stare* e os dois são aceites na língua portuguesa. Por conseguinte, esta frase não cria nenhum problema à tradução automática, sendo que não precisa de ser controlada.

(18) Non credo che si possa cambiare. > Não acredito que possa alterar-se.

²⁷ “In Italian, the equivalent of this verb [*believe*] preferably select the subjunctive, while in the other Romance languages, it is the indicative that most naturally occurs in its complement clause” (Marques, 2010: 148).

²⁸ Como referido em Marques (Ibidem, pp. 145-146), “em português a seleção de um ou outro modo neste tipo de frases depende do grau de crença que se quer expressar. O modo indicativo indica um alto grau de crença enquanto o conjuntivo indica um grau de crença mais baixo”.

²⁹ Embora num registo popular, *encontro* com o mesmo sentido de *trovo* é OK em português.

Esta frase admite o verbo no conjuntivo quer em italiano quer em português, sendo que a negação expressa a não factualidade do enunciado. Embora neste caso o resultado da tradução apresente gramaticalidade, como já referido, o verbo no conjuntivo nem sempre é reconhecido pelo sistema para ser traduzido corretamente na língua de *output*.

Veja-se a frase seguinte:

(19) Non penso che David pratichi abbastanza sport. > *Não penso que David pratica bastante desporto.

Não sendo possível a obtenção do verbo no conjuntivo, o resultado da tradução é agramatical. Para poder resolver este tipo de problema, a negação tem que ser colocada na completiva em vez que na principal:

LC:

(19) a. Penso che David non pratici abbastanza sport. > Penso que David não pratica bastante desporto.

Na tradução da frase (19) a., conforme o grau de crença do locutor, o indicativo é aceite, pelo que não apresenta agramaticalidade. Sendo que (19) e (19) a. têm o mesmo significado, é possível estabelecer uma regra para a linguagem controlada. No entanto, nem sempre se pode adotar este procedimento, pois em certos tipos de frase a subida da negação pode mudar o sentido da frase.³⁰ O controlo com a colocação da negação na completiva é possível apenas com verbos epistémicos.

FRASES COMPLETIVAS NÃO FINITAS.

Em italiano, como já referido, a completiva não finita (*implicita*) tem o verbo no infinitivo e é introduzida pela preposição *di*. Adicionalmente, os sujeitos da principal e da subordinada são co-referentes:

Implicita

(20) Giulia crede di stare bene.

‘Giulia crê de estar bem.’

“Giulia crê que está bem.”

Esplicita

(20) a. *Giulia crede che [Giulia] sta bene.

‘Giulia crê que está bem.’

³⁰ Ver Marques (2010).

Nas frases completivas objetivas italianas cujos sujeitos (da principal e da subordinada) são co-referentes, ocorre apenas a construção *implicita*, sendo a sua forma *esplicita* agramatical.

Em português a forma não finita também é muito usada, mas em tipos de completivas diversas do italiano, como por exemplo, as completivas sujeito. A frase não é introduzida por complementador e o verbo é conjugado no infinitivo pessoal:

Non finita

(21) É importante (nós) estarmos juntos.

A frase na forma finita também é aceite na língua portuguesa:

Finita

(21) a. É importante que (nós) estejamos juntos.

A forma não finita em italiano não é traduzida corretamente para português, pelo que é preciso controlar a linguagem com a forma finita:

(20) Giulia crede di stare bene. > *Julie acredita está bem.

LC:

(20) a. Giulia crede che lei sta bene. > Julie acredita que está bem.

Além da estrutura finita, também é melhor usar o verbo *pensar* (tradução de *pensare*) porque, enquanto o verbo italiano *credere* exprime seja uma crença seja uma opinião, o verbo português *acreditar* exprime mais uma crença do que uma opinião. Normalmente em português expressa-se uma opinião com o verbo *achar* ou *pensar*, enquanto o verbo *acreditar* é usado menos frequentemente.

Segue um outro exemplo com o verbo *credere*, ao qual será aplicado o mesmo tipo de controlo:

(21) Credo di non essere in grado di terminare il lavoro. > *Credo não ser em condições³¹ da terminar trabalho.

³¹ *Essere in grado* é uma expressão da língua italiana que em português pode ser traduzida para “estar em condições” ou, mais simplesmente, “ser capaz” entendido como ter habilidade. O sistema não identifica a oposição ser/estar, portanto é preciso substituir esta expressão por *essere capace*.

Neste caso, o tipo de problema é de natureza diferente: o verbo *credere* não é traduzido e “ser em condições” não é uma expressão aceite neste contexto. Contudo, a estrutura da forma não finita é traduzida corretamente, pelo facto de o verbo da completiva não precisar de flexão:

(21) a. Penso di non essere capace di terminare il lavoro. > Penso não ser capaz de terminar o trabalho.

Sendo apenas um caso particular e não válido para todos os tipos de completivas não finitas – como demonstrado em (20) – o controlo será aplicado com a frase finita, igualmente adequado:

(21) b. Penso che io non sono capace di terminare il lavoro. > Penso que não sou capaz de terminar o trabalho.

4.4 FRASES TEMPORAIS

A subordinada temporal pode expressar uma relação temporal de anterioridade, de posterioridade ou de simultaneidade relativamente ao tempo do evento da frase principal. Em italiano, em termos gerais, para expressar simultaneidade e anterioridade, usa-se o verbo no indicativo e, para expressar posterioridade, usa-se o verbo no conjuntivo. Em português, diferentemente do que acontece em italiano, pode usar-se o conjuntivo também para expressar a anterioridade. Nestes casos, as frases expressam um evento futuro na subordinada e, não havendo certeza de que tal evento aconteça, o modo selecionado é o conjuntivo, no tempo futuro:

“The Future Subjunctive turned out to be the verbal category par excellence for marking future indeterminacy. This is the case for relative clauses whose denotation, i.e. the composition of the denotational set referred to by the modified NP, is still unknown at the moment of utterance” (Becker, 2009: 193).

Em tais contextos, em italiano usa-se o indicativo presente ou futuro. Consequentemente, é preciso ter em conta que as frases que exprimem anterioridade podem dar problemas à tradução automática. As frases temporais introduzidas por *quando* podem expressar quer simultaneidade quer anterioridade e, por conseguinte, em português podem selecionar seja o indicativo seja o conjuntivo na subordinada. Comparam-se este tipo de temporais em italiano e em português:

SIMULTANEIDADE:

(22) Quando (“todas as vezes que”) viene a casa mia, porta sempre i fiori. (**presente do indicativo**; presente do indicativo)

(22) a. Quando (“todas as vezes que”) vem a minha casa, leva sempre flores. (**presente do indicativo**; presente do indicativo)

ANTERIORIDADE:

(23) Chiamami quando (“appena”³²) sei pronto. (imperativo; **presente do indicativo**)

(23) a. Liga-me quando (“logo que”) estiveres pronto. (imperativo; **futuro do conjuntivo**)

Veja-se a maneira em que o sistema seleciona o modo nas frases que expressam anterioridade:

(24) Quando finisci di lavorare, vieni a prendermi. > *Quando terminas trabalhar, vens tomar-me.

O sistema não efetua uma tradução correta, provavelmente pelo facto de *quando* poder expressar também uma relação de simultaneidade. Por conseguinte, substitui-se *quando* por *appena*:

LC:

(24) a. Appena finisci di lavorare, viene³³ a prendermi. > *Logo que terminares trabalhar, vem tomar-me.³⁴

Como se pode constatar, o SystraNet efetua uma tradução correta quando a frase que expressa um evento futuro é introduzida por *appena*, pois seleciona o modo conjuntivo. Assim sendo, a substituição de *quando* por *appena* em frases como (24) faz com que o sistema traduza corretamente, pois o verbo é conjugado no futuro do conjuntivo.

³² *appena* (tal como *logo que*) difere de *quando* na medida em que remete para um intervalo de tempo imediatamente subsequente ao do evento da principal, o que não é o caso de *quando*.

³³ Controlo do imperativo afirmativo para a segunda pessoa do singular (*viene > vieni*).

³⁴ O verbo *prendere*, usado por muitas expressões em italiano e com muitas traduções diferentes em português, muitas vezes, é traduzido de maneira errada pelo SystraNet e não é possível aplicar nenhum controlo da linguagem. Além disso, a frase apresenta agramaticalidade também por não ter *de* entre o verbo *terminar* e *trabalhar*.

Veja-se também a frase (25):

(25) Chiamami quando sei pronto. > *Chame mim quando estás pronto.

Da mesma maneira, substitui-se *quando* por *appena*:

LC:

(25) a. Chiama appena sei pronto. > *Chama mal está pronto.

É importante também a realização do sujeito para evitar que o sistema traduza por estruturas agramaticais, como no caso de (25) a.:

LC:

(25) b. Chiama appena tu sei pronto. > Chama logo que estiveres pronto.

Outro detalhe que é preciso ter em conta no controlo da linguagem, no caso de a frase ter o verbo na segunda pessoa do plural, é a substituição pela terceira pessoa do plural, para que a tradução seja correta em português³⁵:

(26) Chiama appena siete pronti. > *Chama logo que estiver prontos.

LC:

(26) a. Chiama appena loro sono pronti. > Chama logo que estiverem prontos.

Em italiano, assim como em português, as frases que exprimem posterioridade, geralmente, requerem o verbo no modo conjuntivo (presente ou passado), sendo geralmente introduzidas por uma expressão adverbial que integra *che*:

(27) È meglio uscire prima che inizi a piovere.

(27) a. É melhor sair antes que comece a chover.

Todavia, em português também é possível expressar o mesmo valor temporal com o verbo no infinitivo pessoal na subordinada:

(27) b. É melhor sair antes de começar³⁶ a chover.

Em italiano a forma não finita para as temporais não é aceite:

³⁵ Questão que será aprofundada no capítulo 7.

³⁶ Embora “chover” não tenha sujeito argumental, a frase tem sujeito expletivo nulo. Contudo, dado que se trata de terceira pessoa do singular, o infinitivo pessoal não apresenta marcas morfológicas distintivas do infinitivo não pessoal.

(27) c. *È meglio uscire prima di iniziare a piovere.

O problema do SystraNet, como já foi verificado em muitos exemplos, é o reconhecimento do conjuntivo em italiano para traduzi-lo corretamente para português. Entretanto, o sistema também não reconhece *prima che* como *antes que* e traduz por *primeiro*, o que não acontece com *prima di*.³⁷

Assim, a solução será substituir a frase finita pela forma não finita, no caso de frases com o verbo com sujeito expletivo nulo, como no caso de (27):

(27) È meglio uscire prima che inizi a piovere. > *É melhor sair primeiro que inicia chover.

LC:

(27) c. È meglio uscire prima di iniziare a piovere. > É melhor sair antes de começar a chover.

Tal acontece porque, neste caso, o verbo no infinitivo italiano tem a mesma forma do infinitivo da terceira pessoa do singular em português. Por esta razão, este tipo de controlo é possível quando – e só quando – a tradução em português envolve a primeira ou a terceira pessoa do singular, cuja forma é idêntica à do infinitivo não flexionado:

(28) Andiamo via prima che lui ritorni. > *Vamos primeiro que retorna.

LC:

(28) a. Andiamo via prima di lui ritornare > Vamos antes de ele retornar

No caso da terceira pessoa do plural, por exemplo, já não é possível aplicar esta regra porque em português o infinitivo flexionado (infinitivo + *-em*) não tem a mesma forma do infinitivo não flexionado:

(29) Andiamo via prima che ritornino. > *Vamos primeiro que retornam.

LC:

(29) a. Andiamo via prima di loro ritornare. > *Vamos antes de eles retornar.

(Correta: *Vamos antes de eles **retornarem**.*)

Neste último caso, como demonstrado, não há hipótese de fazer com que o sistema traduza pelo infinitivo flexionado em português.

³⁷ Fenómeno que será aprofundado no capítulo 7.

Contudo, pode tentar-se um controlo para que resulte a frase finita. Como se pode notar, o sistema reconhece que *prima che/di* é uma locução conjuncional/preposicional só com a preposição *di*; por conseguinte, colocando esta preposição entre *prima* e *che*, o sistema entende que se trata da locução conjuncional *antes que*.

LC:

(29) b. Andiamo via prima di che loro ritornino. > *Vamos antes que retornam.

De qualquer modo, uma vez mais, o sistema não reconhece o verbo no conjuntivo, pelo que o controlo não é bem sucedido.

4.5 FRASES CONDICIONAIS

As construções condicionais são compostas por duas proposições, uma das quais expressa a condição necessária para que a outra se verifique. A proposição subordinada condicional, a qual expressa a condição, é, por essa razão, comumente designada “prótase” e a principal, que expressa a consequência da subordinada, designada “apódose”.

Embora as condicionais possam assumir diversos aspetos estruturais, conforme os elementos que as integram nas subordinantes, na nossa língua de partida (italiano) tomam-se em consideração apenas as formas prototípicas introduzidas por *se*, dado que é o conector mais usado nas duas línguas.

Em italiano, assim como em português, existem três tipos de construções condicionais: factual, hipotética, ou contrafactual. Contudo, há diferença de modo entre as duas línguas para expressar os diferentes tipos de condicionais.

Em italiano, as frases condicionais de tipo factual têm o verbo no indicativo, tanto na principal como na subordinada. Muitas frases condicionais com ambos os verbos no indicativo podem também expressar uma possibilidade no futuro e ser, por conseguinte, condicionais de tipo hipotético:

(30) Se mangi molti dolci, ingrassi. (condicional de tipo factual, assume-se que o conteúdo das duas proposições é verdadeiro)

(31) Se continua/continuerá a piovere, resterò a casa. (possibilidade futura, tipo de condicional hipotética porque a prótase remete para uma probabilidade de valor

“médio” de realização do evento, que, contudo, e como é natural, ainda não se verificou)

A frase (30), por conseguinte, forma-se com os mesmos modos e tempos verbais nas duas línguas. A frase (31), em contrapartida, ao ser traduzida para português, exigirá um futuro do conjuntivo na prótase, sendo que exprime uma modalidade não factual:

(30) a. Se comes muitos doces, engordas. (indicativo/indicativo)

(31) a. Se continuar a chover, fico em casa. (conjuntivo/indicativo)

O SystraNet reconhece as construções condicionais hipotéticas e contrafactuais e traduz corretamente os verbos *quer* no condicional *quer* no conjuntivo. Em seguida são apresentados exemplos que demonstram este fenómeno.

Hipotéticas:

- **IT: futuro ou presente do indicativo na prótase e futuro do indicativo na apódose;**

PT: futuro do conjuntivo na prótase e futuro do indicativo na apódose.³⁸

(32) Se continua/continuerà a piovere, resterò in casa. > Se continuar a chover, permanecerei em casa.

(33) Se arriviamo/arriveremo in tempo, andremo a giocare a tennis insieme.³⁹ > Se chegarmos a tempo, iremos jogar ao ténis juntos.

- **IT-PT: imperfeito do conjuntivo na prótase e condicional simples na apódose**

(34) Se andassi a Milano, passerei a salutare Maria.⁴⁰ > Se fosse à Milão, passaria cumprimentar Marie.

(35) Se venissi con me, ti divertiresti molto. > Se viesses comigo, divertir-se-ias muito.

³⁸ A referência às línguas feita nesta secção justifica-se com a necessidade de contrastar a variação do modo nas frases condicionais das duas línguas e também os respetivos resultados da tradução automática, para demonstrar o bom desempenho do sistema relativamente à seleção do modo.

³⁹ Para uma correta tradução, nesta frase deve ser controlada a locução adverbial *in tempo* com *a tempo* e deve omitir-se a preposição *a* antes de *tennis*.

⁴⁰ Para uma correta tradução, nesta frase deve ser controlada a preposição *a* (antes de *Milano*) substituindo-a por *la*.

Contrafactuais:

- **Presente; IT-PT: imperfeito conjuntivo na prótase e condicional simples na apódose**

(36) Andrei più spesso all'estero, se conoscessi le lingue. > Iria mais frequentemente no estrangeiro⁴¹, se conhecesse as línguas.

(37) Se avessi soldi, gli comprerei un bel regalo. > Se tivesse dinheiro, comprar-lhe-ia um bonito presente.

- **Passado; IT-PT: mais do que perfeito do conjuntivo prótase e condicional composto na apódose**

(38) Se fossi stato più gentile, avresti ottenuto quello che volevi. > Se tivesse sido mais agradável, terias obtido \emptyset ⁴² que querias.

(39) Se fosse arrivato prima delle 6, avrebbe visto suo padre. > Se tivesse chegado antes de 6 horas, teria visto \emptyset ⁴³ seu pai.

Sendo que por um lado o SystraNet está programado de maneira a que consiga reconhecer as condicionais hipotéticas e contrafactuais, por outro lado não reconhece os tipos de condicionais factuais. Em italiano as factuais têm a mesma estrutura das hipotéticas, enquanto em português as hipotéticas requerem o futuro do conjuntivo. Assim, uma frase condicional factual, que requer o indicativo em ambas as línguas, é traduzida por uma frase hipotética com o conjuntivo futuro:

(30) Se mangi molti dolci, ingrassi.⁴⁴ > *Se comeres muitas sobremesas, você engrossados.

LC:

(30) a. Se mangi molti dolci, diventi grasso. > Se comeres muitas sobremesas, ficas gordo.

Como é demonstrado, o SystraNet seleciona o conjuntivo futuro na tradução de um verbo no presente do indicativo depois da conjunção condicional *se*. Para controlar a linguagem nas frases factuais, será preciso omitir esta preposição e alterar de alguma

⁴¹ Problema preposicional (tratado na secção 7.2): “ir ao estrangeiro” seria a expressão correta, sendo que *ir* é um verbo de movimento. Não se chegou a uma forma eficaz de controlo.

⁴² Falta o item pronominal *o*. Não se chegou a uma forma eficaz de controlo.

⁴³ Falta o artigo antes de *seu*. Não se chegou a uma forma eficaz de controlo.

⁴⁴ Dado que o sistema não reconhece o verbo *ingrassare* (“engordar”) conjugado na segunda pessoa do singular, substituímo-lo pela perífrase *diventare grasso* (“ficar gordo”).

maneira a ordem da frase, para que o sistema não traduza pelo conjuntivo futuro mas sim pelo indicativo presente.

A maneira mais eficaz para controlar a linguagem será conjugar o verbo da subordinada condicional no gerúndio presente e especificar o sujeito na frase principal:

LC:

(30) b. Mangiando molti dolci, tu diventi grasso. > Comendo muitas sobremesas, ficas gordo.

(40) Se l'acqua arriva alla temperatura di 100 gradi, entra in ebollizione. > *Se a água chegar à temperatura de 100 graus, entra em ebulição.

LC:

(40) a. Arrivando alla temperatura di 100 gradi, l'acqua entra in ebollizione. > Chegando à temperatura de 100 graus, a água entra em ebulição.

Contudo, ao aplicar este controlo, o resultado de tradução não é muito natural na expressão de uma frase condicional em português. No entanto, desta maneira, é possível evitar a tradução para o futuro do conjuntivo e, em consequência, denotar a factualidade deste tipo de orações para poder diferenciá-las das condicionais hipotéticas.

4.6 FRASES CONCESSIVAS

As frases concessivas expressam a falta do efeito que poderia ou deveria resultar de uma determinada causa. Em outras palavras, a realização do evento da concessiva não impede a realização do evento da subordinante, contrariamente ao expetável.

Em italiano, tendo em consideração apenas as concessivas factuais e as hipotéticas, – não considerando, portanto, as contrafactuais – a seleção do modo é determinada pelo conector que as introduz: *benché* e *sebbene* requerem o verbo no conjuntivo presente; *anche se* requer o verbo no indicativo presente. Em português, a maioria dos conectores que introduzem uma concessiva requer o conjuntivo e alguns requerem o infinitivo pessoal.

O SystraNet reconhece as concessivas contrafactuais, pois têm a mesma estrutura verbal nas duas línguas (conjuntivo mais-que-perfeito/condicional composto):

(41) Anche se fosse arrivato in tempo⁴⁵, non avrebbe visto lo spettacolo. > Ainda que tivesse chegado a tempo, não teria visto o espetáculo.

Relativamente às concessivas factuais e hipotéticas, o sistema dificilmente consegue traduzir adequadamente os modos verbais. De facto, o SystraNet traduz corretamente os verbos no conjuntivo só das subordinadas concessivas introduzidas por *benché*:

(41) Benché abbia molta fame, non mangerò. > Embora tenha muita fome, não comerei.

Anche se é traduzido por *ainda que*, conector que requer o verbo no conjuntivo. Todavia, o SystraNet não traduz corretamente o verbo porque o deixa no modo indicativo. Por conseguinte, nas frases concessivas introduzidas por *anche se* é preciso substituir este conector por *benché*, que requer o verbo no conjuntivo:

(42) Anche se ho ragione, non voglio insistere. > *Ainda que tenho razão, não quero insistir.

LC:

(42) a. Benché io abbia ragione, non voglio insistere. > Embora tenha razão, não quero insistir.

Veja-se também outro exemplo:

(43) Anche se mangi molti dolci, non diventi grasso. > *Ainda que comes muitas sobremesas, não se tornas gordo.

LC:

(43) a. Benché mangi molti dolci, non diventi grasso. > *Embora coma muitas sobremesas, não se tornas gordo.

Uma vez mais, é preciso colocar o sujeito antes do verbo para que o sistema traduza corretamente o verbo na segunda pessoa:

LC:

(43) b. Benché tu mangi molti dolci, tu non diventi grasso > *Embora comas muitas sobremesas, não se tornas gordo.⁴⁶

⁴⁵ Para obter a correta tradução de *in tempo*, deve-se substituir *in* por *a* (ver nota 39).

⁴⁶ Nesta frase é preciso controlar a linguagem por causa da má tradução do verbo *diventare*, porque em português pode expressar-se de formas diferentes. *Tornar-se* é apropriado neste contexto, mas não é

LC:

(43) c. Benché tu mangi molti dolci, tu non sei grasso > Embora comas muitas sobremesas, não és gordo.⁴⁷

A mesma substituição pode ser aplicada às concessivas introduzidas por *sebbene*. De facto, este conector requer um verbo no conjuntivo, o seu correspondente em português – *embora*:

(44) Sebbene possa piovere, usciremo lo stesso. > *Embora pode chover, sairemos o mesmo.

O sistema considera o advérbio *embora* como a tradução de *sebbene*, mas não traduz o verbo no modo conjuntivo, razão pela qual é preciso substituí-lo por *benché*:

LC:

(44) a. Benché possa piovere, usciremo lo stesso. > ?? Embora possa chover, sairemos o mesmo.

LC:

(44) b. Benché possa piovere, usciremo nella stessa.⁴⁸ > Embora possa chover, sairemos na mesma.

4.7 CONCLUSÃO

A variação de modo de italiano para português causa vários tipos de problemas à tradução automática, não só devidos às diferenças que existem nas duas línguas, mas também resultantes de insuficiências do sistema de natureza diversa.

Após várias tentativas e, apesar das falhas do sistema, encontraram-se soluções para o controlo da linguagem na maioria dos casos que são aqui tratados. Em alguns destes, o texto de *input* tem de ser objeto de degradação para que o *output* seja aceitável (como, por exemplo, no caso das construções completivas). Todavia, em outros casos, é preciso apenas proceder a pequenas alterações sem que tal acarreta agramaticalidade, como no

conjugado corretamente; a única possibilidade seria substituir esse verbo por outro com valor semântico aproximando e pragmaticamente equivalente, por exemplo *ser*.

⁴⁷ A tradução correta deveria ser com o verbo *estar*. Mesmo assim, infere-se que a expectável mudança de estado não ocorre.

⁴⁸ Controlo da linguagem com a substituição de *lo stesso* (“na mesma”) por *nella stessa*.

caso das concessivas (*sebbene* > *benché*) ou das temporais com relação de anterioridade (*quando* > *appena*). Em contrapartida, verificaram-se casos em que, no contexto deste trabalho, não foi possível encontrar uma solução para o controlo da linguagem (por exemplo, nalguns casos de imperativo), por razões de inadequação de especificações do sistema e eventualmente por inoperatividade na iteração das suas componentes.

No capítulo a seguir, são tratadas as variações do par linguístico italiano-português no âmbito da modalidade. Cabe salientar que, tendo em conta a forte ligação entre modo e modalidade, muitos problemas tratados aqui apresentam-se também no tratamento da modalidade, pelo que alguns deles podem ser resolvidos com as regras propostas neste capítulo.

5. MODALIDADE

O conceito de modalidade é normalmente relacionado com a conceção lógica ou linguística que se lhe atribui, sendo que, durante séculos, a questão tem sido um objeto de estudo bastante controverso. Do ponto de vista linguístico, a modalidade pode ser definida como “a gramaticalização de atitudes e opiniões dos falantes” (Oliveira, 2003: 245). A maneira mais comum para expressar conceitos modais é através de verbos modais – nomeadamente, *dever* e *poder* – mas também através de outros tipos de verbos, advérbios, adjetivos ou expressões epistémicas.

Existem dois tipos de modalidade, nomeadamente epistémica e deôntica. A modalidade epistémica envolve as noções de possibilidade e necessidade, enquanto a modalidade deôntica envolve as noções de obrigação e permissão. Em muitas línguas, significados deônticos e epistémicos podem ser expressos pelas mesmas formas, como é o caso do italiano e do português. Contudo, poder-se-á constatar que as duas línguas apresentam diferenças com as quais é preciso lidar.

Neste capítulo, são tomados em consideração os verbos modais (ou seja, a forma mais comum para expressar a modalidade, quer deôntica, quer epistémica) e as variações destes do italiano para português. Além disso, consideram-se também algumas expressões epistémicas italianas que criam problemas à tradução pela sua falta de correspondência em português e propõem-se soluções para um controlo da linguagem adequado.

5.1 VERBOS MODAIS

5.1.1 *DOVERE* > *DEVER*

O verbo modal *dovere* em italiano, da mesma forma que *dever* em português, pode ser usado quer num contexto epistémico quer num contexto deôntico. Todavia, entre as duas línguas há muitas diferenças de uso destes verbos.

Em italiano, o verbo *dovere* pode ser usado em muitos contextos diferentes.

(1) Franco deve uscire per comprare le sigarette.

‘Franco deve sair para comprar os cigarros.’

Nesta frase, o verbo *dovere* tem um valor deôntico, sendo que pode ser substituído por *é necessario che*, por ser uma necessidade que o sujeito tem. Com efeito, *deve* denota obrigação (ter de sair) gerada por uma causa (comprar os cigarros) (Parisi *et al.*, 1975; *apud* Pietrandrea, 2002: 84). Por conseguinte, a frase correspondente em português tem de ser “reforçada” através de *ter de*, auxiliar de modalidade mais usado na língua portuguesa para as frases deônticas ou para reforçar o grau de probabilidade nas frases epistémicas. Assim sendo, a tradução para português será a seguinte:

(1) a. Franco tem de sair para comprar os cigarros.

O modal da frase (1) diferencia-se do modal das frases que seguem:

(2) I riflettori di questi giorni non gli devono aver fatto troppo piacere, perché hanno spostato l’attenzione. (epistémico factual-inferencial)

‘Os projetores destes dias não lhe devem ter agradado muito, porque lhe desviaram a atenção.’

(3) Giovanni deve essere uscito perché non vedo il suo cappotto. (estritamente epistémico)

‘Giovanni deve ter saído porque não vejo o casaco dele.’

(4) Ci deve essere un bavaglino nella borsa, prendilo! (epistémico factual)

‘Deve haver um bibe na mala, pega nele!’

Em todas estas frases (2-4), o valor do modal *dovere* é epistémico, mas em cada frase tem significado ligeiramente diferente. Em (2), o verbo *dovere* introduz uma conclusão baseada em conhecimentos objetivos, pelo que é uma frase de tipo factual-inferencial. Neste caso, em português a modalidade epistémica expressa-se principalmente com o mesmo modal, *dever*:

(2) I riflettori di questi giorni non gli devono aver fatto troppo piacere, perché hanno spostato l’attenzione.

(2) a. Os projetores destes dias não lhe devem ter agradado muito, porque lhe desviaram a atenção.

A frase (3) é estritamente epistémica, no sentido em que se trata de uma hipótese baseada na convicção do falante, pelo que a frase não exprime certeza, mas um elevado grau de probabilidade. A frase correspondente em português tem a mesma interpretação:

(3) Giovanni deve essere uscito perché non vedo il suo cappotto.

(3) a. Giovanni deve ter saído porque não vejo o casaco dele.

A frase (4) tem um elemento que determina o grau epistémico máximo; isto é, o imperativo (*prendilo!*). O facto de o falante usar este modo verbal implica convicção relativamente ao conteúdo da frase modalizada, que, desta forma, assume valor factual. Em português, como há um grau de probabilidade muito maior do que na frase (3), será preciso usar *ter de*, já que “reforça a convicção” (Costa Campos, 1998: 129). Contudo, há outros elementos que determinam a factualidade de um enunciado, como por exemplo, algumas expressões que explicitam a atitude/opinião do locutor, como “tenho a certeza”, “sem dúvida”, etc.

(5) Questo quadro deve essere di Picasso, non ho dubbi.

(5) a. Este quadro tem de ser de Picasso, não tenho dúvidas.

A factualidade da frase (5), por exemplo, é determinada pela expressão *non ho dubbi* (“não tenho dúvidas”), assim como o é pelo imperativo na frase (4):

(4) Ci deve essere un bavaglino nella borsa, prendilo!

(4) a. Tem de haver um bibe na mala, encontra-o!⁴⁹

Por conseguinte, para algumas destas frases, será preciso o controlo da linguagem. Para que o sistema traduza corretamente, a semântica da frase italiana deve ter o verbo *avere* (“ter”) + a preposição *di* (“de”):

(1) Franco deve uscire per comprare le sigarette.

LC:

(1) b. Franco ha di uscire per comprare le sigarette. > Franco tem de sair para comprar os cigarros.

(5) Questo quadro deve essere di Picasso, non ho dubbi.

LC:

⁴⁹ O imperativo da frase italiana, *prendilo!* (“pega nele!”), deve ser traduzido de uma maneira diferente para português, ou seja “encontra-o!”. Embora os dois verbos tenham um significado diferente, em português, neste tipo de contexto, não é natural usar *pegar* ou qualquer sinónimo. Esta questão pragmática deve ser tida em conta no controlo da linguagem.

(5) b. Questo quadro ha di essere di Picasso, io non ho dubbi. > Este quadro tem de ser de Picasso, não tenho dúvidas.

(4) Ci deve essere un bavaglino nella borsa, prendilo!

LC:

(4) b. Ha di esserci un bavaglino nella borsa, prendilo!⁵⁰ > *Tem ser um babeiro na bolsa, toma!

Sendo que na tradução para português tem que resultar o verbo *encontrar*, será preciso modificar o verbo *prendere* e substituí-lo com *trovare*:

LC:

(4) c. Ha di esistere un bavaglino nella borsa, trovalo! > *Tem de existir um babeiro na bolsa, encontrada!

Também surge um problema de tradução do verbo no imperativo. Como já explicado anteriormente, a frase imperativa pode ser controlada com a transformação do verbo na terceira pessoa do singular do presente do indicativo, especificando o pronome pessoal:

LC:

(4) d. Ha di esistere un bavaglino nella borsa, lui lo trova! > Tem de existir um babeiro na bolsa, encontra-o!

Em italiano, normalmente, o verbo *dovere* tende a dar um significado mais factual comparado com as outras línguas, enquanto em português a factualidade é “reforçada” através de *ter de*.

Contudo, o verbo *dever* em português é muito usado em qualquer tipo de frase epistémica. O uso de *ter de* na frase epistémica/factual reforça apenas o grau de probabilidade, mas pode usar-se também o verbo *dever*, sem alteração do significado. O que é mais difícil encontrar é o verbo *dever* na frase deôntica, razão pela qual é melhor controlar a linguagem com *ter de*.

O verbo *dever* em português, contrariamente a *dovere* em italiano, é um verbo defetivo, pelo que não se conjuga no pretérito perfeito simples. Em italiano, sendo que *dovere* existe em todos os seus tempos verbais, ao submeter uma frase italiana com este verbo

⁵⁰ Para controlar a linguagem é preciso mudar a forma verbal *esserci* porque causa problemas. Esta estrutura impessoal não é reconhecida pelo sistema. O seu equivalente pode ser o verbo *existir*.

em tempos passados como o *passato remoto* e o *passato prossimo*, o sistema encontrará dificuldades na tradução:

(6) Maria dovette uscire prima per non perdere l'aereo. > *Marie dut sair primeiro para não perder o avião.

(6) a. Maria è dovuta uscire prima per non perdere l'aereo. > *Marie é devida sair primeiro para não perder o avião.

(6) b. Maria ha dovuto uscire⁵¹ prima per non perdere l'aereo. > *Marie deveu sair primeiro para não perder o avião.

Face a este tipo de problemas, a solução será, uma vez mais, usar *ter de*.

Dado em italiano o tempo *passato remoto* – equivalente gramatical do pretérito perfeito simples – não ser muito usado⁵², o SystraNet nem sempre o reconhece:

io ebbi	*mim ter
tu avesti	*você ter
lui ebbe	teve
noi avemmo	*nós eúmes ⁵³
voi aveste	? tinha
loro ebbero	tiveram

Tabela 1. *Passato remoto* do verbo *avere* e resultados de tradução do SystraNet.

Assim sendo, para controlar a linguagem, será preciso conjugar o verbo no *passato prossimo*, para que o sistema traduza corretamente:

LC:

(6) c. Maria ha avuto di uscire prima per non perdere l'aereo. > Marie teve de sair *primeiro⁵⁴ para não perder o avião.

⁵¹ Nas frases (6) a. e (6) b. usam-se ambos os auxiliares *essere* e *avere* para demonstrar que o sistema não reconhece, em nenhum dos casos, a defetividade do verbo *dever* em português e, por conseguinte, traduz por *deveu*.

⁵² Este tempo verbal usa-se mais no Sul de Itália do que no Norte. Contudo, tem um uso diferente do pretérito perfeito do português, pois em italiano este tempo verbal usa-se para contar acontecimentos considerados como cumpridos num passado psicologicamente longínquo e sem relações com o presente. Este problema será tratado no capítulo 6.

⁵³ O sistema utiliza a língua francesa como língua ponte.

⁵⁴ Um problema que se pode notar nesta frase é a tradução do advérbio *prima* neste contexto. De facto, *prima* é traduzido para *primeiro*, mas neste caso não é a tradução correta porque o que queremos

LC:

(6) d. Maria ha avuto di uscire con antecedenza per non perdere l'aereo. > Marie teve de sair com antecedência para não perder o avião.

Da mesma maneira, no caso de ocorrência numa frase com o verbo *dovere* no *passato prossimo*, é preciso substituir este verbo por *avere di*, conjugado também no *passato prossimo*:

(7) Ho dovuto insistere ma alla fine l'ho convinto. > *Devi insistir mas no fim convenci-o.

LC:

(7) a. Ho avuto di insistere ma alla fine l'ho convinto. > Tive de insistir mas finalmente convenci-o.

5.1.2 POTERE > PODER

O verbo italiano *potere* parece ser radicado no domínio deôntico, muito mais do que *dovere*. Em particular, o uso de *potere* no modo indicativo dificilmente tem valor epistémico e, quando tem, há uma sobreposição com o valor deôntico:

DEÔNTICO:

(8) Franco può uscire. \cong È possibile che Franco esca.
'Franco pode sair.' \cong 'É possível que Franco saia.'

EPISTÉMICO:

(9) Gianni può essersi fermato. \cong È possibile che Gianni si sia fermato.⁵⁵
'Gianni pode ter parado.' \cong 'É possível que Gianni tenha parado.'

Havendo uma sobreposição entre a interpretação deôntica e a epistémica, o verbo italiano *potere* remete para um domínio mais deôntico do que epistémico. De facto, na frase epistémica com o verbo *potere*, normalmente este verbo é acompanhado por

expressar na frase é "com antecedência". A tradução mais apropriada desta frase obtém-se substituindo *primeiro* pela locução adverbial *com antecedência*.

⁵⁵ Adaptado a Pietrandrea (2002: 78-79).

outros verbos, formando desta maneira expressões epistémicas, como, por exemplo, *può darsi* ou *può essere*. Este tipo de expressões existe também na língua portuguesa.

Em particular, a forma *pode ser* tem uma distribuição semelhante à de *può essere*. No caso de uma proposta ou de uma pergunta direta, *pode ser* pode ter um grau mais alto de probabilidade:

(10) - Queres ir ao cinema esta noite?

- Pode ser.

A resposta “pode ser”, mesmo não tendo a mesma assertividade de “sim, quero”, pode ser interpretada como uma resposta positiva, na medida em que exprime a não oposição à proposta. Em contrapartida, se este complexo verbal forma uma frase não factual com o verbo no conjuntivo, pode ser equiparado à expressão modal epistémica italiana *può essere*:

(11) Pode ser que o Andrea esteja em casa.

(11) a. Può essere che Andrea sia a casa sua.

Uma outra expressão epistémica que integra o verbo *potere*, muito usada em italiano, é *può darsi*, mas a sua forma em português “pode dar-se” é de uso muito mais limitado. Uma expressão parecida e mais corrente no uso da língua é “pode dar-se o caso de”.

Para a tradução automática, as duas expressões *può darsi* e *può essere* poderiam ser controladas de diferentes maneiras, mas todas as formas modais não factuais deveriam ter o conjuntivo na subordinada – ex. “é possível que”, “pode ser que”, etc. + conjuntivo presente. Como já explicado anteriormente, o sistema nem sempre reconhece o verbo no conjuntivo, pelo que no controlo da linguagem será preciso procurar uma forma adverbial que expresse o mesmo grau de epistemicidade e que seja independente do modo na subordinada:

(12) Può darsi che sia un caso. > *Pode dar-se que seja um caso.

LC:

(12) a. Possibilmente è un caso. > Possivelmente é um caso.

No caso de (12), o sistema reconhece *sia* como conjuntivo presente e traduz corretamente por *seja*. Contudo, como já demonstrado anteriormente, *sia* é normalmente

reconhecido como conjuntivo pelo sistema, mas a forma do conjuntivo de muitos outros verbos não o é. Por conseguinte, em todos os casos, será melhor substituir *può darsi* por uma forma adverbial que não selecione o conjuntivo:

(13) Può darsi che Luisa venga alla festa. > *Pode dar-se que Luisa vem à festa.

LC:

(13) a. Possibilmente Luisa viene alla festa. > Possivelmente Luisa vem à festa.

Como se pode observar na frase (13), o verbo *venire* no conjuntivo não é reconhecido pelo sistema. O controlo com o advérbio *possibilmente* – (13) a. – melhora o seu desempenho.

A mesma substituição pode ser aplicada a *può essere*, quando implica a presença do conjuntivo:

(11) b. *Può essere* che Andrea sia a casa sua. > **Pode ser* que Andrea está à casa o seu.

LC:

(11) c. *Possibilmente* Andrea è dentro casa.⁵⁶ > *Possivelmente* Andrea está em casa.

5.2 EXPRESSÕES MODAIS EPISTÉMICAS

Muitas vezes, nas frases epistémicas italianas, encontram-se algumas expressões modais que perderam o seu significado etimológico para assumir um significado epistémico. Algumas delas, como explicado na secção anterior, são formadas pelo verbo *potere* na terceira pessoa do singular mais *essere* ou *darsi*. Todavia, existem também outras expressões que são completamente independentes de qualquer verbo modal.

Seguem alguns exemplos:

(15) Si vede che aveva molta fame.

‘Vê-se que tinha muita fome.’

Neste caso, o verbo *vedere* é conjugado na forma não flexionada (*si vede*) e não indica um acto de percepção, mas sim uma expressão epistémica.

(16) Capace che sei stanco.

‘Capaz que estás cansado.’

⁵⁶ Construção adaptada de modo a permitir “em casa”, posto que o sistema não traduz corretamente a construção italiana *dentro casa*.

Da mesma maneira, neste contexto, *capace* não exprime uma habilidade.

(17) Mi sa che Giovanni non torna a casa oggi.

‘Sabe-me que Giovanni não volta a casa hoje.’

Neste caso, também o verbo *sapere* impessoal não indica conhecimento.

Estas conjeturas epistémicas distinguem-se das expressas por formas verbais comuns por serem fixas, não permitindo nenhum tipo de flexão (Pietrandrea, 2002: 45):

(15) a. *Si vedeva che aveva molta fame.

‘Veia-se que tinha muita fome.’

(16) a. *Capaci che sono stanchi.

‘Capazes que estão cansados.’

(17) a. *Mi sapeva che Giovanni non tornava a casa oggi.

‘Sabia-me que Giovanni não voltava a casa hoje.’

Na língua portuguesa, estas expressões não existem. Só existe uma forma parecida ao *capace che*, ou seja *ser capaz*, na sua forma pessoal (Oliveira, 2001: 171 *apud* Pietrandrea, 2002: 46):

(16) Capace che sei stanco. → Ele é capaz de estar cansado.

(18) Capace che arrivi domani. → Ele é capaz de chegar amanhã.

Na tradução de *si vede che*, será preciso utilizar outras expressões ou modais epistémicos que existam na língua portuguesa, por exemplo *dever*:

(15) Si vede che aveva molta fame. → Deve ter tido muita fome.

(19) Si vede che ha fatto tardi e si vergognava a dirlo. → Deve ter chegado tarde e tinha vergonha de dizê-lo.

Assim sendo, a equivalência expressa pelo verbo modal *dever* do português indica o facto de a frase ter um significado epistémico/factual, pressuposto pela expressão *si vede che*.

Veja-se o resultado da tradução automática de (15):

(15) Si vede che aveva molta fame. > Vê-se que tinha muita fome.

Como se pode observar, o resultado é uma frase gramaticalmente correta, mas o significado é diferente do da frase italiana. De facto, a frase portuguesa remete mais

para um ato de percepção e não tanto para uma avaliação epistêmica. Assim sendo, a escolha do verbo *dovere* para o controlo da linguagem é sem dúvida a melhor:

LC:

(15) a. Deve aver avuto molta fame. > Deve ter tido muita fome.

O mesmo controlo pode ser aplicado à outra frase:

(19) Si vede che ha fatto tardi e si vergognava a dirlo. > *Vê-se que tem feito atrasado e tinha-se vergonha a dizê-lo.

LC:

(19) a. Deve aver fatto tardi e si vergognava a dirlo. > *Deve ter feito atrasado e tinha-se vergonha a dizê-lo.

Neste caso, há vários problemas de tradução que provocam uma construção da frase particularmente mal formada⁵⁷, pelo que deve ser controlada mais uma vez:

LC:

(19) b. Deve essere arrivato tardi e vergognava di dirlo. > Deve ter chegado atrasado e tinha vergonha de dizê-lo.

Uma outra expressão epistêmica muito usada em italiano é *mi sa che*, seguida pelo verbo no indicativo:

(17) Mi sa che Giovanni non torna a casa oggi.

(20) Mi sa che mio fratello va al cinema stasera.

A expressão mais adequada para controlar a expressão *mi sa che* seria *ho l'impressione che*, mas isto implica, uma vez mais, a presença do conjuntivo presente, pelo que o sistema poderia não traduzir corretamente:

⁵⁷ Problemas de tradução:

- *Fare tardi* é uma expressão coloquial italiana sem correspondência em português, que é semanticamente equivalente a *arrivare tardi* (“chegar tarde”).

- O sistema não consegue reconhecer o advérbio *tardi* (tarde) e conseqüentemente traduz para o adjetivo. Neste caso não há soluções para controlar a linguagem, porque no dicionário do SystraNet a palavra *tardi* não é reconhecida como equivalente a *tarde*.

- A construção *si vergognava a* + verbo no infinitivo é traduzida literalmente para português. Portanto, para obter a construção certa, temos de usar a construção portuguesa na frase italiana.

(17) a. Ho l'impressione che Giovanni non torni a casa oggi. > *Tenho a impressão que Giovanni não voltas à casa hoje.

O substantivo *impressione* (“impressão”) é o que seleciona o conjuntivo na frase completiva, tanto em italiano como em português. Como já verificado em diversas situações, o sistema tem problemas em reconhecer o conjuntivo presente italiano. Neste caso, *torni* é conjuntivo presente de terceira pessoa do singular, mas pode ser também a forma do indicativo de segunda pessoa do singular. Por conseguinte, o controlo com *ho l'impressione che* não resulta adequado.

Contudo, a forma da frase completiva objetiva com o verbo *pensare* pode ser uma escolha válida para o controlo, sendo que em português é aceite também o verbo no indicativo e pode ter um significado muito próximo à conjectura *mi sa*:

(17) Mi sa che Giovanni non torna a casa oggi.

LC:

(17) b. Penso che Giovanni non torni a casa oggi. > *Penso que Giovanni não voltas à casa hoje.

Como já mencionado na parte das completivas objetivas (secção 4.3), a língua portuguesa, contrariamente à italiana, não requer obrigatoriamente o conjuntivo neste tipo de frases. De facto, em (17) b. apresenta-se o mesmo problema de tradução que ocorre em (17) a., ou seja, a tradução de *torni* como se fosse o indicativo da segunda pessoa do singular. Assim sendo, a única solução será a substituição do conjuntivo pelo indicativo, sendo este aceite na frase portuguesa:

LC:

(17) c. Penso che Giovanni non torna a casa oggi. > *Penso que Giovanni não gira à casa hoje.

Resolvendo os outros problemas de tradução⁵⁸, a frase controlada será a seguinte:

LC:

⁵⁸ A frase tem um problema lexical relacionado com o verbo *tornare*, traduzido para *girar*. A solução seria substituir esse verbo por um sinónimo que seja traduzível corretamente para português, *ritornare* (“retornar”).

Também a expressão “ir a casa” seria mais correto do que “ir à casa” (tradução do SystraNet), portanto controla-se a linguagem com o artigo sem preposição, *la casa*, para que não surja a preposição contraída com o determinante.

(17) c. Penso che Giovanni non ritorna la casa oggi. > Penso que Giovanni não retorna⁵⁹ a casa hoje.

Obviamente, o mesmo controlo pode ser aplicado à outra frase:

(20) a. Mi sa che mio fratello va al cinema stasera.

LC:

(20) b. Penso che mio fratello va al cinema stasera. > Penso que o meu irmão vai ao cinema esta noite.

A expressão italiana *capace che* não permite nenhum tipo de flexão. O seu equivalente português *ser capaz de*, em contrapartida, flexiona-se em todas as pessoas:

(21) Elas são capazes de não vir.

Para traduzir a frase (16), é preciso efetuar um controlo que respeite a estrutura portuguesa. Assim sendo, substitui-se a estrutura não flexionada pela estrutura flexionada, para que o significado do resultado de tradução corresponda ao da frase italiana:

(16) Capace che arrivi domani. > *Capaz que chega amanhã.
'Capaz que chegue amanhã.'

LC:

(16) a. Lui è capace di arrivare domani. > É capaz de chegar amanhã.

A regra pode ser aplicada também a frases com sujeito expletivo nulo:

(22) Capace che non c'era più posto. > *Capaz que mais lá era posto.
'Capaz que não havia mais lugar'

LC:

(22) a. È capace di non esserci più posto. > *É capaz não lá de ser posto mais.

Há vários elementos de confusão para o sistema que impedem bons resultados de tradução.⁶⁰ Assim, a frase controlada terá a seguinte estrutura:

⁵⁹ A tradução para *retornar* não é natural na língua portuguesa. Contudo, foi a única tradução que se consegue obter neste contexto. A questão será aprofundada no capítulo 7.

⁶⁰ A ordem da estrutura faz com que o advérbio *più* seja ligado com *posto* não como se este último fosse um nome, mas sim como se fosse o participio passado do verbo *porre* (“pôr”), ou seja *posto* (“posto”).

LC:

(22) b. È capace di non restare nessun luogo. > É capaz de não permanecer nenhum lugar.

5.3 CONCLUSÃO

Em conclusão, as variações dos verbos modais do italiano para português são de várias ordens e, em todos os casos analisados, foi possível aplicar uma linguagem controlada para resolver os respectivos problemas de tradução automática. Em alguns casos, o controlo da linguagem causa uma degradação do *input* com vista à obtenção de melhores resultados no *output* (como no caso de *ter de*, sendo um verbo português sem correspondência na língua italiana). As expressões epistémicas italianas são um exemplo de não correspondência gramatical na língua portuguesa. De facto, as soluções encontradas para o controlo da linguagem alteram e simplificam ao mesmo tempo a estrutura das frases que contêm este tipo de expressões. Como se pôde constatar no capítulo anterior, um grande problema que tem a ver quer com o modo quer com a modalidade é a questão do conjuntivo (sobretudo no tempo presente), não só pelo que diz respeito ao reconhecimento das formas de alguns verbos pelo SystraNet, mas também em relação ao seu uso em determinadas estruturas, em algumas das quais são reconhecidas e noutras não.

Contudo, é preciso salientar que o tratamento da modalidade não é aqui tratado de maneira totalmente exaustiva, mas apenas ao fim de estabelecer regras para a linguagem controlada para os casos mais problemáticos que prejudicam a eficiência do sistema para a tradução automática do par italiano-português.

Portanto, é melhor tirar esse advérbio e substituir a frase pelo verbo *rimanere* e o adjetivo *nessun* e *luogo* em vez que *posto*.

6. ASPETO

O aspeto tem a ver com a estrutura interna do evento descrito pela frase. Há que distinguir “aspeto” de “tempo” e de *aktionsart*. O tempo localiza os eventos relativamente ao momento da fala ou a um tempo de referência expresso na frase. Em português, assim como em italiano, o tempo e o aspeto podem ser expressos pelo mesmo morfema na flexão verbal. O aspeto é fundamentalmente gramatical porque é dependente das formas gramaticais e, em consequência, é ligado aos tempos verbais. O *aktionsart*, em contrapartida, é de natureza lexical, sendo ligado ao significado do lexema verbal (Oliveira, 2003: 129-203).

Basicamente, distinguem-se dois tipos de aspeto: perfeito e imperfeito. O aspeto perfeito implica a conclusão do evento, enquanto o aspeto imperfeito exprime o evento “no seu acontecer”. O aspeto imperfeito pode ser dividido em progressivo, contínuo e habitual, conforme o tipo de intervalo de tempo tomado em consideração.

Neste capítulo são tratadas as diferenças na expressão do aspeto entre o italiano e o português. Tais diferenças ocorrem sobretudo nos tempos verbais que podem ou não marcar um certo grau de perfetividade, como no caso do *passato prossimo* e do *passato remoto*.

Primeiro trata-se dos tempos *passato prossimo* e *passato remoto* e de como se comportam do ponto de vista aspetual, comparando-os com os equivalentes portugueses, nomeadamente o pretérito perfeito composto o pretérito perfeito e simples. A partir dessa análise, tenta-se encontrar soluções para estabelecer regras para o controlo do aspeto perfeito e imperfeito/iterativo nestes tempos verbais.

Em seguida, também são tratadas as estruturas perifrásticas progressivas, as quais têm características comuns nas duas línguas, mas também algumas diferenças, sobretudo do ponto de vista aspetual. Analisa-se como se portam estas perífrases na tradução automática e tenta-se encontrar soluções para eventuais problemáticas.

6.1 *PASSATO PROSSIMO* E *PASSATO REMOTO*

Em relação à estrutura gramatical destes tempos verbais, o *passato prossimo* pode ser equiparado ao pretérito perfeito composto e o *passato remoto* ao pretérito perfeito

simples. Em contrapartida, em termos aspetuais, há algumas diferenças importantes que devem ser tidas em consideração.

Em português, o pretérito perfeito composto “não marca perfeitividade mas uma duração que tem início no passado e perdura no presente” (Oliveira, 2003: 142).

Em italiano, pelo contrário, o correspondente *passato prossimo* pode expressar um evento acontecido no passado que tem (ou não) alguma relação com o presente. Assim, este tempo tem uma dupla valência aspetual, pois pode ser interpretado seja no sentido perfeito, seja no sentido imperfeito (Dardano e Trifone, 1995: 355).

É interessante notar a especificidade do uso do pretérito perfeito composto português, pois diferencia-se dos correspondentes “perfeitos” das línguas românicas. De facto, em português esta forma verbal está associada a uma leitura de iteratividade, normalmente apoiada por expressões verbais como *ultimamente, nos últimos dias, desde* etc.

[...] “[A] major requirement in order to use the Perfect is the durativity or iterativity of the situation, which has to span an interval of time starting in the past and reaching the Speech Time” (Squartini, 1998: 152).

Ao comparar o pretérito perfeito composto português com o *passato prossimo* italiano regista-se uma diferença do ponto de vista aspetual. Enquanto uma frase portuguesa com o verbo no pretérito perfeito composto pode ser traduzida para italiano pelo *passato prossimo*, o inverso nem sempre é possível, por causa da variação de aspeto entre italiano e português:

(1) Ultimamente o João tem lido muitos romances. → Ultimamente João ha letto molti romanzi.

(2) Ieri pomeriggio sono andato in palestra. → *Ontem à tarde tenho ido ao ginásio.
(Correta: *Ontem à tarde fui* ao ginásio)

Em português, o pretérito perfeito composto forma-se com o auxiliar *ter* seguido pelo verbo no particípio passado. A estrutura gramatical do *passato prossimo* italiano é ligeiramente diferente, sendo que este tempo pode ser formado seja com o auxiliar *avere* (“ter”) seja com o auxiliar *essere* (“ser”), conforme a transitividade do verbo no particípio passado⁶¹.

⁶¹ O verbo *avere* usa-se geralmente para verbos ativos intransitivos, o verbo *essere* usa-se para verbos transitivos, impessoais e reflexivos. Contudo, a escolha dos dois verbos auxiliares com os verbos intransitivos não segue critérios constantes e regulares.

O tempo *passato prossimo* pode ser usado para descrever um evento que ainda se está a verificar no presente:

(3) Due giorni fa ho preso una brutta influenza.

‘Dois dias faz tenho apanhado uma má gripe.’

“Há dois dias apanhei uma forte gripe.”

Ou para descrever um evento cujos efeitos continuam no presente:

(4) Ho imparato l’inglese durante un soggiorno di studio negli Stati Uniti.

‘Tenho aprendido inglês durante uma estada de estudo nos Estados Unidos.’

“Aprendi inglês durante um período de estudo nos Estados Unidos.”

(5) Marco é nato il 21 settembre 1987.

‘Marco tem nascido o 21 setembro 1987.’

“Marco nasceu no dia 21 de setembro de 1987.”

Na gramática italiana, o *passato prossimo* está associado terminologicamente ao *passato remoto*, pois o elemento de diversificação entre os dois tempos verbais, normalmente, é a colocação mais ou menos próxima no eixo temporal. Todavia, como descrito por Dardano e Trifone (1995: 355), o que mais conta na escolha dos dois tempos verbais é a maior ou menor “atualidade psicológica” em relação ao evento descrito, sendo que nem sempre é considerada a proximidade do evento relativamente ao presente:

(6) Dio ha creato il mondo.

‘Deus tem criado o mundo.’

“Deus criou o mundo.”

(7) L’invenzione della scrittura ha rivoluzionato i rapporti tra popoli.

‘A invenção da escritura tem revolucionado as relações entre os povos.’

“A invenção da escritura revolucionou as relações entre os povos.”

O *passato remoto* italiano pode equiparar-se ao pretérito perfeito português, pois os dois tempos verbais expressam um evento concluído no passado, independentemente da distância do tempo em relação ao presente e das suas eventuais relações com o presente.

Na língua contemporânea, muitas vezes, o *passato remoto* é substituído pelo *passato prossimo*, sobretudo em algumas regiões do Norte de Itália:

(8) L'anno scorso sono andato in Brasile.

'No ano passado tenho ido em Brasil.'

"No ano passado fui ao Brasil."

As razões são, basicamente, de natureza psico-linguística, pois o falante, na maioria dos casos, quer aproximar os factos ao momento da narração. Pelo contrário, no Sul de Itália, usa-se muito mais o *passato remoto* do que o *passato prossimo* para expressar a perfeitividade, mesmo quando se fala de acontecimentos muito próximos do presente (Dardano e Trifone, 1995: 355):

(9) Arrivai un quarto d'ora fa.

'Cheguei há um quarto de hora.'

Em ambos os tempos verbais (*passato prossimo* e *passato remoto*), o SystraNet traduz por verbos conjugados no pretérito perfeito simples, o tempo da perfeitividade da língua portuguesa:

(10) Ho imparato l'inglese durante un soggiorno di studio negli Stati Uniti. > Aprendi Inglês durante uma estada de estudo nos Estados Unidos.

(11) Mio padre arrivò in Italia nel 1956. > *O meu pai chegou na Itália em 1956.

LC:

(11) a. Mio padre arrivò a Italia nel 1956.⁶² > O meu pai chegou à Itália em 1956.

Contudo, como já foi mencionado anteriormente, na secção 5.1.1, muitos verbos conjugados no *passato remoto* não são reconhecidos pelo SystraNet e, por conseguinte, tem de controlar-se a linguagem para que o sistema traduza corretamente.

Vejamos alguns exemplos:

(12) L'anno scorso viaggiai per l'America. > *O ano passado mim voyagai para a América.

(13) Manzoni nacque nel 1785.⁶³ > *Manzoni nacquit em 1785.

O SystraNet parece passar por outras línguas (línguas ponte) para tentar traduzir o *passato remoto* destes verbos.

⁶² Neste caso, a frase precisa de ser controlada no que respeita a preposição *em*, uma vez que *chegar* seleciona a preposição *a*.

⁶³ Usa-se o *passato prossimo* para indicar o nascimento de uma pessoa/personagem ainda vivo, mas é obrigatório usar o *passato remoto* para indicá-lo de uma pessoa falecida (Dardano e Trifone, 1995: 355).

Com efeito, como já visto no caso do verbo *dovere*, o sistema não reconhece as formas de *passato remoto* de alguns verbos. Estão nesse caso os dois verbos em causa – *viaggiare* e *nascere* –, que não têm flexão regular no *passato remoto*, não traduzindo o SystraNet nenhuma forma senão a terceira pessoa do plural de *viaggiare*:

io nacqui	*mim nacquis
tu nascesti	*você nacquis
lui nacque	*ele nacquit
noi nascemmo	*nos nacquîmes
voi nasceste	*nascia
loro nacquero	*nacquirent

Tabela 1. *Passato remoto* do verbo *nascere* e resultados de tradução do SystraNet.

io viaggiai	*mim voyagai
tu viaggiasti	*você voyagas
lui viaggiò	*ele voyaga
noi viaggiammo	*nos voyagâmes
voi viaggiaste	? viajava
loro viaggiarono	*viajaram

Tabela 2. *Passato remoto* do verbo *viaggiare* e resultados de tradução do SystraNet.

Para evitar este tipo de traduções, e tendo em conta que o sistema traduz também o *passato prossimo* pelo pretérito perfeito simples, é preciso conjugar o verbo no *passato prossimo* sempre que a frase apresente perfeitividade:

(12) L'anno scorso viaggiai per l'America.

LC:

(12) a. L'anno scorso ho viaggiato per l'America. > O ano passado viajei para a América.

(13) Manzoni nacque nel 1785.

LC:

(13) a. Manzoni é nato nel 1785. > Manzoni nasceu em 1785.

(11) Mio padre arrivò in Italia nel 1956.

LC:

(11) b. Mio padre è arrivato a Italia nel 1956. > O meu pai chegou à Itália em 1956.

Como o sistema traduz sempre pelo pretérito perfeito simples, será difícil fazer com que traduza pelo pretérito perfeito composto, ou seja, quando a frase expressa iteratividade:

(14) Ultimamente ho mangiato troppo. > ? Ultimamente comi demasiado.

(15) Negli ultimi giorni Maria è arrivata sempre puntuale. > ? Nos últimos dias Marie chegou sempre pontual.

Todavia, para resolver o problema com o controlo da linguagem, pode-se ter em consideração o verbo *tenere*, sendo uma opção para fazer com que a tradução para português resulte *ter*. Em termos semânticos, *tenere* quer dizer “manter, conservar, guardar”, pelo que não tem nenhuma função auxiliar na língua italiana. Com efeito, o uso deste verbo em função de auxiliar faz com que a frase italiana fique completamente agramatical. Contudo, o que interessa é a obtenção de um *output* adequado e, por isso, o controlo da linguagem é efetuado através de uma degradação do *input*:

LC:

(14) a. Ultimamente tengo mangiato troppo. > Ultimamente tenho comido demasiado.

LC:

(15) a. Negli ultimi giorni Maria tiene arrivato sempre puntuale. > Nos últimos dias Marie tem chegado sempre pontual.

6.2 CONSTRUÇÕES PERIFRÁSTICAS PROGRESSIVAS

A perífrase italiana com o verbo *stare*, contrariamente ao que acontece nas outras línguas românicas, refere-se apenas a situações imperfetivas progressivas e não aceita funções estritamente durativas (Squartini, 1998: 127).

O verbo *stare* pode formar dois tipos de perífrases, assim como o verbo *estar* em português:

1. *stare* + gerúndio → *estar* + gerúndio

2. *stare a* + infinitivo → *estar a* + infinitivo

stare + gerúndio é a estrutura perifrástica mais comum e mais utilizada na língua italiana. Esta perífrase tem a função de descrever eventos que estão a decorrer num determinado intervalo de tempo, seja este presente (16), passado (17) ou futuro (18):

(16) In questo momento, Marco sta lavorando.

‘Neste momento, Marco está trabalhando.’

(17) Quando Elena è entrata in casa, Maria stava mangiando una mela.

‘Quando Elena entrou em casa, Maria estava comendo uma maçã.’

(18) Quando avrai 30 anni, starai ancora vivendo coi tuoi genitori.

‘Quando tiveres 30 anos, ainda estarás vivendo com os teus pais.’

stare a + infinitivo também tem uma função progressiva, mas o seu uso é bastante limitado, pois é circunscrito ao sentido de “passar certo tempo a fazer qualquer coisa”, muitas vezes com um sentido negativo de “desperdiçar tempo” (Squartini, 1998: 128).

(19) Non starò a raccontare tutta la mia storia.

‘Não estarei a contar toda a minha história.’

Num contexto perfeitivo, apenas *stare a* + infinitivo pode ser aceite, também com a combinação do advérbio *sempre* ou locuções adverbiais como *per tutto il tempo*, *per tutta la sera*:

(20) Paolo è stato a parlare con Luca *per tutta la sera*.

‘Paolo esteve a falar com Luca por toda a noite.’

Em contrapartida, a construção perifrástica *stare* + gerúndio nunca é aceite em contextos perfeitivos. *stare a* + infinitivo pode também ser usado para processos durativos, também com a combinação do advérbio *sempre* (Squartini, 1998: 130):

(21) Paolo sta sempre a parlare con Luca.

‘Paolo está sempre a falar com Luca.’

Do ponto de vista aspetual, as duas estruturas perifrásticas são formas diferentes no domínio da progressividade: *stare* + gerúndio indica progressividade em contextos imperfeitivos, enquanto *stare a* + infinitivo é usado em estruturas progressivas em contextos iterativos e perfeitivos (Squartini, 1998: 132).

Em português, as duas estruturas correspondentes são as variedades europeia e brasileira: em português europeu a forma não marcada é estar a + infinitivo enquanto em português brasileiro e em algumas variantes do português europeu se usa estar + gerúndio (Squartini, 1998: 112). Contrariamente ao que acontece em italiano, as duas estruturas perifrásticas podem ser conjugadas nos tempos perfeitos, pelo que podem ter também características perfeitivas:

(22) Ontem estive a trabalhar todo o dia. (variedade europeia)

(23) Ontem estive trabalhando todo o dia. (variedade brasileira e variante dialetal alentejana)

Contudo, tais diferenças aspetuais, a causarem problemas, será na tradução português → italiano, e não em sentido inverso – o relevante neste trabalho –, já que em italiano apenas stare a + infinitivo pode ocorrer – e ainda assim raramente – num contexto perfeitivo.

Ao submeter à tradução automática frases com construções perifrásticas, veja-se o desempenho do sistema:

(16) In questo momento, Marco sta lavorando. > Neste momento, Marco trabalha.

(17) Quando Elena è entrata in casa, Maria stava mangiando una mela. > Quando Elena entrou em casa, Marie comia uma maçã.

(18) Quando avrai 30 anni, starai ancora vivendo coi tuoi genitori. > Quando tiveres 30 anos, ainda viverás com os teus pais.

Como é possível constatar, em caso de perífrases progressivas, esteja o auxiliar no presente, passado ou futuro, o sistema não consegue traduzi-las pelas frases correspondentes em português. O sistema traduz estes tipos de construções, respetivamente, por estruturas com os verbos nos tempos presente, imperfeito e futuro simples do indicativo. Por conseguinte, do ponto de vista temporal, os resultados de tradução das frases (16), (17) e (18) colocam corretamente a situação no eixo temporal. Por outro lado, do ponto de vista aspetual, estes tempos verbais não denotam a progressividade das estruturas perifrásticas.

Em particular, no caso do resultado de tradução da frase (16), deve ser analisado também um fator relacionado com a diferença em termos aspetuais entre o presente do

indicativo italiano e o presente do indicativo português. No caso específico da frase (16), o verbo traduzido no presente é aceite porque o acto de trabalhar pode ser entendido como ocorrendo num intervalo de tempo que envolve o momento da fala, mas que se estende ao futuro e eventualmente ao passado próximo⁶⁴. Contudo, o discurso torna-se diferente se se considera, por exemplo, a seguinte frase:

(24) Marco sta mangiando. > Marc come.

Neste caso, embora a frase seja gramatical, a tradução não é aceitável pelo valor aspetual do tempo presente em português, diferente do italiano.

Como descrito em Oliveira (2003: 144), “do ponto de vista aspetual, o Presente do Indicativo apresenta a interessante característica de só ser um tempo presente, pelo menos parcialmente sobreposto ao tempo da enunciação, com estados.” “Com eventos, está restringido a relatos directos e ao uso de enunciados performativos” (Ibidem, pp. 154). Para descrever processos, “a leitura preferencial [do Presente do Indicativo] é de estado habitual, construído com base numa ocorrência indeterminada de eventos do mesmo tipo que têm lugar num intervalo de tempo não delimitado, mas que inclui o tempo da enunciação” (Ibidem, pp. 144). No caso de processos culminados, este tempo raramente é usado, senão no caso de ser acompanhado por adverbiais que expressam quantificação ou duração de tempo (Ibidem, pp. 144).

Por conseguinte, sendo que este tempo tem uma leitura de habitualidade quando se trata de processos, o resultado da frase (24) não se pode considerar uma tradução correta, pois expressa um estado habitual enquanto deveria expressar a progressividade do processo.

Dado o desempenho do sistema na tradução das perífrases progressivas, pode hipoteticar-se que haja alguma regra estabelecida em relação às progressivas italianas, cujos correspondentes verbos portugueses são estabelecidos conforme o tempo do auxiliar *stare* (sta lavorando > trabalha; stava mangiando > comia; etc.).

Contudo, há maneira de controlar a linguagem. Com a substituição da estrutura stare + infinitivo por stare per + infinitivo, o sistema traduz pela perífrase estar a + infinitivo:

(16) Attualmente, Marco sta lavorando nell’impresa di suo padre.

⁶⁴ A frase (16) é ambígua: o facto de Marco estar a trabalhar pode querer dizer que atualmente tem um emprego ou que neste preciso instante está no local de trabalho. O significado depende do contexto, mas o resultado da tradução (Marco trabalha) só pode ser aceite no primeiro caso.

LC:

(16) a. Attualmente, Marco sta per lavorare nell'impresa di suo padre. > Atualmente, Marco está a trabalhar na empresa de seu pai.

(17) Quando Elena è entrata in casa, Maria stava mangiando una mela.

LC:

(17) a. Quando Elena è entrata in casa, Maria stava per mangiare una mela. > Quando Elena entrou em casa, Marie estava a comer uma maçã.

(18) Quando avrai 30 anni, starai ancora vivendo coi tuoi genitori.

LC:

(18) a. Quando avrai 30 anni, starai per vivere ancora⁶⁵ coi tuoi genitori. > Quando tiveres 30 anos, ainda estarás a viver com os teus pais.

(24) Marco sta mangiando.

LC:

(24) a. Marco sta per mangiare. > Marc está a comer.

A perífrase italiana usada para o controlo da linguagem, stare per + infinitivo, expressa a eminência de realização do evento denotado pelo verbo (infinitivo). Esta estrutura, por conseguinte, não se pode considerar equivalente a stare + gerúndio, pois não expressa progressividade na língua de *input*. Nesse caso, estamos perante uma frase controlada que não causa degradação gramatical do *input*, mas que lhe altera completamente o significado. Contudo, este facto não tem qualquer relevância, já que permite a tradução desejada para a frase original, sendo de resto, a única forma de obter estar a + infinitivo, na forma adequada.

Consideramos agora a outra perífrase progressiva italiana, stare a + infinitivo.

Segue o exemplo com o verbo no presente:

(19) Non starò a raccontare tutta la mia storia. > Não contarei qualquer minha história.⁶⁶

LC:

⁶⁵ Mudança de posição do advérbio *ancora* para deixar junta a perífrase na língua de *input* e permitir uma correta tradução.

⁶⁶ Para fazer com que a tradução resulte correta, é preciso trocar a ordem de determinante-nome, ou seja substituir *mia storia* por *storia mia*.

(19) a. Non starò a raccontare tutta la storia mia. > Não contarei toda a história minha.

Veja-se também o exemplo com o verbo no passado:

(20) Paolo è stato a parlare con Luca per tutta la sera. > *Paolo foi a falar com Luca por toda a noite.⁶⁷

LC:

(20) a. Paolo stette a parlare con Luca per tutta la sera. > Paolo falou com Luca por toda a noite.⁶⁸

Mesmo com este tipo de perífrases progressivas, o sistema não traduz as correspondências para português. Tal como acontece com as frases consideradas precedentemente, também as frases (19) a. e (20) a. não são agramaticais, mas do ponto de vista aspetual não correspondem ao expresso na língua de *input*. Uma vez mais, a situação está colocada no tempo certo, mas não é expressa a progressividade do aspeto, sobretudo no caso da (20), cuja perífrase é traduzida pelo verbo no pretérito perfeito, expressando, desta maneira, uma ideia de perfeitividade⁶⁹.

A linguagem é controlada através da substituição da perífrase stare a + infinitivo por stare per + infinitivo:

(19) Non starò a raccontare tutta la storia mia.

LC:

(19) b. Non starò per raccontare tutta la storia mia. > *Não serei em comboio de contar toda a história minha.

Neste caso, o problema de tradução é devido à negação expressa por *non*, sendo que sem esta, o sistema traduz corretamente:

LC:

(19) c. ø Starò per raccontare tutta la storia mia. > Estarei a contar toda a história minha.

⁶⁷ O sistema não reconhece a construção perifrástica com este verbo no *passato prossimo*, pelo que se substitui pelo *passato remoto*.

⁶⁸ As frases (19) e (20) são controladas respetivamente com (19) a. e (20) a., por razões externas ao tópico em causa.

⁶⁹ A progressividade do resultado da frase (20) a. é inferida só pela expressão temporal “por toda a noite”.

LC:

(20) b. Paolo stette per parlare con Luca durante tutta la sera. > Paolo esteve a falar com Luca durante toda a noite.

Em consequência, nas frases negativas não se pode usar a mesma regra que para as afirmativas, pois o SystraNet não consegue reconhecer a perífrase antecedida pela negação. Todavia, como foi possível verificar nas frases afirmativas (19) c. e (20) b., o sistema traduz corretamente estas perífrases a partir de stare per + infinitivo, razão pela qual é possível a criação de uma regra.

Contudo, um problema que se poderá verificar na leitura do *input* neste tipo de linguagem controlada, será confundir a função de iminência própria da construção *stare per + infinitivo* com a função de progressividade pretendida com o controlo destas estruturas. Para desambiguar este tipo de frase, será preciso criar uma regra para as frases que denotam iminência de realização do evento. Contudo, este problema não será aqui tratado, dado não caber nos objetivos deste trabalho. Todavia, não se exclui o seu tratamento em possíveis trabalhos futuros.

6.3 CONCLUSÃO

Em conclusão, pode dizer-se que o aspeto perfeito e o imperfeito são expressos de maneira bastante diferente nas duas línguas. As variações aspetuais de italiano e português ocorrem principalmente nos chamados “tempos perfeitos”, nomeadamente pretérito perfeito simples e composto.

No caso do *passato remoto*, ocorrem problemas relacionados com o desempenho do sistema, pois os verbos conjugados neste tempo verbal, na sua maioria, não são reconhecidos, pelo que são traduzidos de forma errada. Apesar de o sistema processar corretamente alguns verbos de uso mais frequente (como *chegar*), o controlo da linguagem com o *passato prossimo* é necessário para permitir ao sistema a tradução pelo pretérito perfeito simples português, o tempo correspondente em português para expressar perfeitividade.

Em contrapartida, para que na língua de *output* fique a ideia de imperfetividade, é necessário controlar a linguagem permitindo uma degradação da língua de *input* para obter a tradução do verbo no pretérito perfeito composto.

Com respeito às construções perifrásticas progressivas, apesar de terem poucas diferenças a nível aspetual em italiano e em português, o SystraNet não oferece qualquer resultado que incorpore a ideia de progressividade presente na língua de *input*. Todavia, foi encontrada uma outra perífrase italiana que permite uma tradução correta, embora seja alterado o seu significado na língua de *input*. Mesmo assim, apesar dos problemas que ocorrem na tradução de frases negativas, foi possível a criação de regras para o controlo da linguagem das estruturas perifrásticas progressivas.

7. OUTROS FENÓMENOS A REQUERER CONTROLO

Neste capítulo são tratados alguns fenómenos problemáticos que, embora não caíam nos objetivos centrais deste trabalho, foram identificados nos vários exemplos analisados no âmbito da persecução desses objetivos, i.e., do estabelecimento de regras para controlo dos *inputs* em matéria de modo, modalidade e aspeto.

Os fenómenos tratados neste capítulo são, por conseguinte, na sua maioria, exemplos de dificuldades que se colocaram no processo de controlo das matérias em questão. Os vários exemplos são retomados e renumerados em todas as secções deste capítulo. Os problemas classificados e tratados incluem não apenas os causados pelas diferenças linguísticas entre italiano e português, mas também os relacionados com o mau desempenho do sistema, independentemente de tais diferenças.

Os vários tipos de problemas captados surgem aqui agrupados sob as seguintes designações gerais: aspetos lexicais e aspetos sintáticos. Em alguns casos, o problema pode ser resolvido através da linguagem controlada, enquanto em outros não. Nos casos em que o controlo for possível, estabelecem-se regras para o efeito, sendo apresentados e explicados os vários testes efetuados para melhorar o resultado da tradução automática do SystraNet.

7.1 ASPETOS LEXICAIS

TORNARE > VOLTAR. A frase (1) a., abaixo, é controlada a partir da frase (1), sendo o verbo conjugado no indicativo para facilitar a tradução, embora tal acarrete agramaticalidade:

(1) *Mi sa che Giovanni non torna a casa oggi.* > *Me sabe que Giovanni não gira à casa hoje.

LC:

(1) a. *Penso che Giovanni non torna a casa oggi.* > *Penso que Giovanni não gira à casa hoje.

O resultado da tradução tem um problema lexical relacionado com o verbo *tornare*, traduzido por *girar*, provavelmente porque foi usado o inglês como língua ponte (*tornare* > *turn* > *girar*). O sistema não regista o significado *voltar/regressar* para os verbos italianos testados e, assim sendo, a única solução possível será substituir o verbo *tornare* por um sinónimo que seja traduzível corretamente para português.

Após várias tentativas, foi escolhido o verbo *ritornare* porque é o único que pode ser facilmente processável pelo sistema. Há que referir que o verbo mais adequado seria certamente *voltar*, mas o sistema parece não traduzir corretamente todas as formas deste verbo a partir do correspondente italiano *tornare*, razão pela qual é melhor utilizar *ritornare* e, assim, obter o *output* com *retornar*.

LC:

(1) b. Penso che Giovanni non ritorna a casa oggi. > Penso que Giovanni não retorna à casa hoje.

PRENDERE > TOMAR. Na frase (2) foi substituído o advérbio *quando* por *appena*, para permitir a ocorrência do futuro do conjuntivo na tradução da frase temporal:

(2) Quando finisci di lavorare, vieni a prendermi. > *Quando terminas trabalhar, vens tomar-me.

LC:

(2) a. Appena finisci di lavorare, viene a prendermi. > *Logo que terminares trabalhar, vem tomar-me.

O resultado de tradução desta frase é agramatical por causa da falta de *de* entre *terminares* e *trabalhar* e porque o verbo *prendere*, na expressão *venire a prendermi* (“vir buscar-me”), é traduzido por *tomar*, verbo muito utilizado na língua portuguesa e com muitos significados diferentes.

Estamos perante um caso de ambiguidade lexical, pois o verbo na língua de *input* tem mais do que uma correspondência na língua de *output*. No caso, o SystraNet escolhe a correspondência de um dos verbos mais produtivos. O problema é que o verbo correspondente ao da expressão *venire a prendermi* não é *tomar*, mas *buscar*. No dicionário do SystraNet, como se pode verificar, há várias aceções para este verbo:

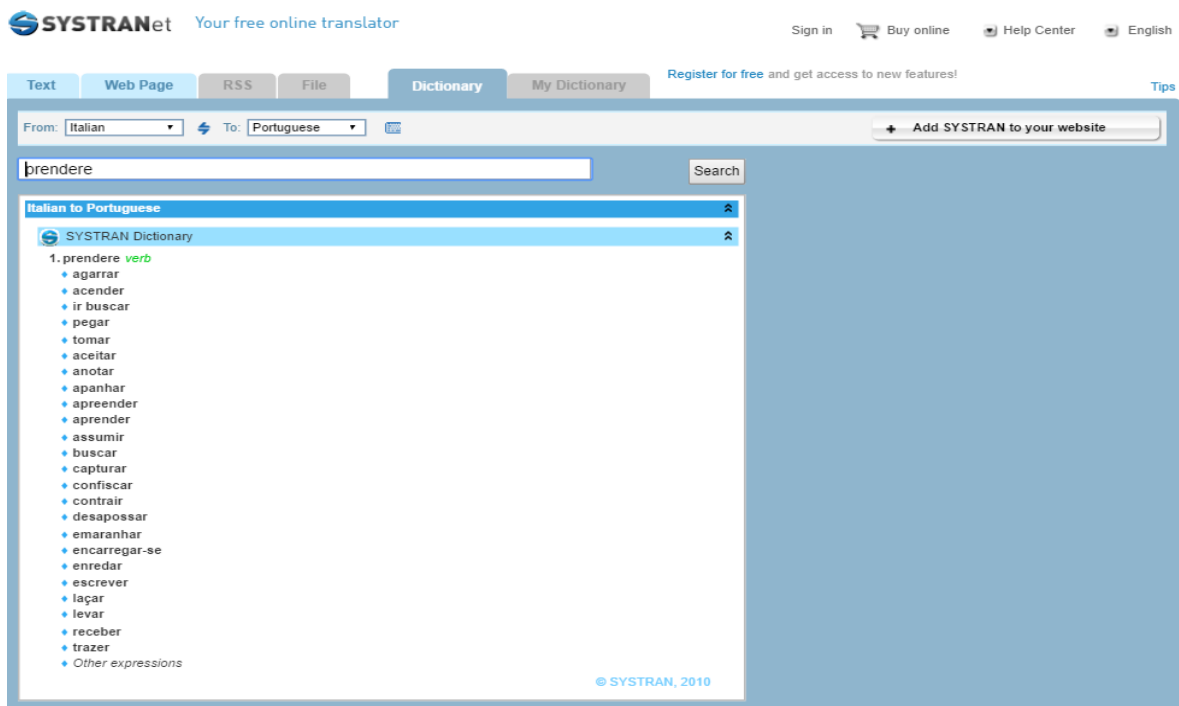


Figura 1. Correspondências de *prendere* no dicionário do SystraNet.

Entre as várias correspondências, está *buscar*, bem como *ir buscar*, como deveria ser traduzido no caso da frase (2). O verbo *buscar* seria o equivalente semântico do italiano *cercare*, mas também de *procurar*. Assim sendo, mesmo substituindo o verbo *prendere* por *cercare*, não é selecionado o verbo *buscar*, mas sim *procurar*, que não é a tradução adequada neste contexto:

LC:

(2) b. Appena finisci di lavorare, viene a cercarmi. > *Logo que terminares trabalhar, vem procurar-me.

Por conseguinte, não há maneira de criar uma regra para obter o verbo *buscar* no *output*. Assim, o problema ligado à polissemia do verbo *prendere*, neste contexto, não pode resolver-se.

Para solucionar o problema respeitante à falta da preposição *de*, pode substituir-se o verbo pelo nome correspondente:

LC:

(2) c. Appena finisci il lavoro, viene a prendermi. > ? Logo que terminares o trabalho, vem tomar-me.

LO STESSO > NA MESMA. A frase (3) a. foi controlada apenas com a substituição do advérbio *benché*, para que a tradução resulte com verbo no conjuntivo:

(3) Sebbene possa piovere, usciremo lo stesso. > *Embora pode chover, sairemos o mesmo.

LC:

(3) a. Benché possa piovere, usciremo lo stesso. > *Embora possa chover, sairemos o mesmo.

Na frase (3) a., a expressão *lo stesso* é usada como “reforço” da asserção de que a ocorrência do evento denotado pela frase subordinante não é posta em causa pelo eventual obstáculo expresso na concessiva. A expressão correspondente em português pode ser *na mesma* ou *igualmente*, mas a tradução do sistema é literal (*o mesmo*), sendo o resultado agramatical. Para controlar o problema, podem ter-se em consideração duas opções: 1. substituir *lo stesso* por *nella stessa* (“na mesma”); 2. substituir *lo stesso* pelo advérbio *ugualmente* (“igualmente”), de idêntico significado.

LC:

(3) b. Benché possa piovere, usciremo ugualmente. > Embora possa chover, sairemos igualmente.

LC:

(3) c. Benché possa piovere, usciremo nella stessa. > Embora possa chover, sairemos na mesma.

Para a criação da regra, a opção mais adequada é a segunda, como no controlo efetuado em (3) c. A escolha é devida ao facto de permitir um melhor resultado no *output*, embora induza degradação da boa-formação do *input*, que é irrelevante, dada a orientação do controlo estritamente para a tradução automática.

INGRASSARE > ENGORDAR. A frase (4) a., controlada a partir de (4), contém um verbo de mudança de estado, *ingrassare* (“engordar”), que não é reconhecido pelo sistema:

(4) Se mangi molti dolci, ingrassi. > *Se comeres muitas sobremesas, você engrossadas.

LC:

(4) a. Mangiando molti dolci, tu ingrassi. > *Comendo muitas sobremesas, você engrossados.

O sistema traduz o verbo *ingrassare* por *engrossar*, verbo com um significado do mesmo campo semântico de *engordar*, mas totalmente inadequado neste contexto. Além disso, o SystraNet não reconhece a segunda pessoa do singular deste verbo, assim como a primeira do singular e a segunda do plural são traduzidas de forma incorreta:

io ingrasso	*mim engrossados
tu ingrassi	*você engrossados
lui ingrassa	engrossa
noi ingrassiamo	engrossamos
voi ingrassate	engrossa
loro ingrassano	engrossam

Tabela 1. Presente do indicativo do verbo *ingrassare* e resultados de tradução do SystraNet.

Contudo, como acontece em muitos casos, o dicionário do sistema tem uma das correspondências correta, ou seja, *engordar*:

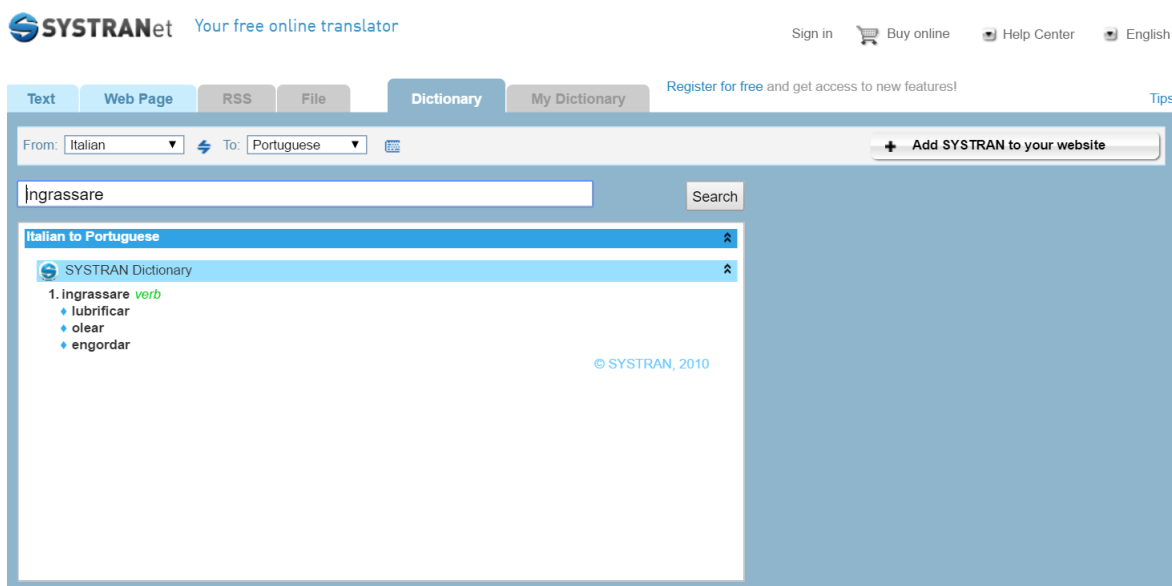


Figura 2. Correspondências de *ingrassare* no dicionário do SystraNet.

Sendo este verbo um verbo de mudança de estado, podemos usar o seu equivalente *diventare* – “ficar”, em português – seguido pelo adjetivo que expressa o estado final

resultante da mudança, ou seja, *grasso* (“gordo”). Por conseguinte, para controlar a linguagem, substitui-se o verbo *ingrassare* por *diventare grasso* (“ficar gordo”):

LC:

(4) b. Mangiando molti dolci, tu diventi grasso. > Comendo muitas sobremesas, ficas gordo.

TU DIVENTI GRASSO / TU NON DIVENTI GRASSO. A frase (5) tem o mesmo problema da (4) a., pois contém o verbo de mudança de estado *ingrassare*, nesse caso antecedido pela negativa *non*:

(5) Benché tu mangi molti dolci, tu non ingrassi. > *Embora comas muitas sobremesas, você engrossados não.

Além dos problemas antes referidos, o sistema coloca a negação depois do verbo. Para controlar a linguagem, portanto, aplica-se a mesma substituição usada na (4) a., ou seja, de *ingrassare* por *diventare grasso*:

LC:

(5) a. Benché tu mangi molti dolci, tu non diventi grasso. > *Embora comas muitas sobremesas, não se tornas gordo.

Neste caso, *tornar-se* é apropriado no contexto, mas o que não é adequado nesta frase é o clítico de terceira pessoa (*se*) antes do verbo na segunda pessoa do singular (*tornas*). O correto, de facto, deveria ser “te tornas gordo”, mas o sujeito *tu* não está em posição de adjacência ao verbo, contrariamente ao que acontece na frase afirmativa. Este parece (a questão requer investigação mais aprofundada) ser o fator de perturbação no que respeita à seleção adequada do clítico:

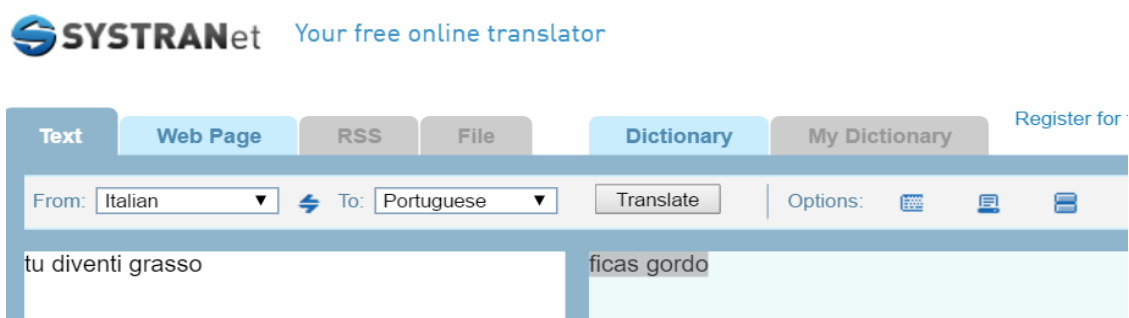


Figura 3. Resultado da tradução de *tu diventi grasso* do SystraNet.

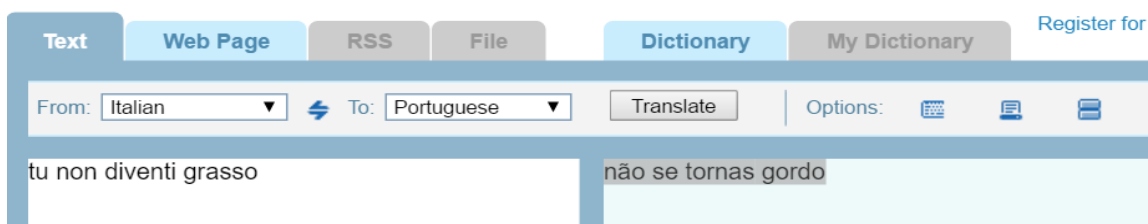


Figura 4. Resultado da tradução de *tu non diventi grasso* do SystraNet.

Por este motivo, não há possibilidade de controlar a linguagem, no contexto desta investigação, para que o verbo em frases negativas seja traduzido corretamente pelo sistema; a única possibilidade será substituir este verbo por outro que permite a inferência do mesmo estado de coisas:

(5) b. Benché tu mangi molti dolci, tu non sei grasso. > Embora comas muitas sobremesas, não és gordo.

Desta maneira, produz-se um melhor resultado de tradução. Embora o verbo devesse ser *estar* pelo facto de se tratar de uma propriedade não inerente, através do verbo *ser* infere-se que a expectável mudança de estado não ocorre.

POSTO > LUGAR. A frase (6) a. foi controlada a partir da (6):

(6) Capace che non c'è più posto. > *Capaz que mais lá é posto.

LC:

(6) a. È capace di non esserci più posto. > *É capaz não lá de ser posto mais.

O resultado de tradução, em ambas as frases, é agramatical e semanticamente ininterpretável. Há vários elementos que criam certa confusão ao sistema e impedem bons resultados de tradução.

O verbo *esserci* cria problemas ao sistema e altera toda a estrutura da frase⁷⁰, pelo que tem de ser substituído. Além disso, a ordem da estrutura italiana faz com que o sistema reconheça o advérbio *più* relacionado com *posto*, como se este último fosse o particípio passado do verbo *porre* (*pôr* em português, *posto*). A palavra *posto*, de facto, tem esta ambiguidade identificada também pelo sistema:

⁷⁰ Os problemas associados a *esserci* são tratados em secção para o efeito, abaixo.

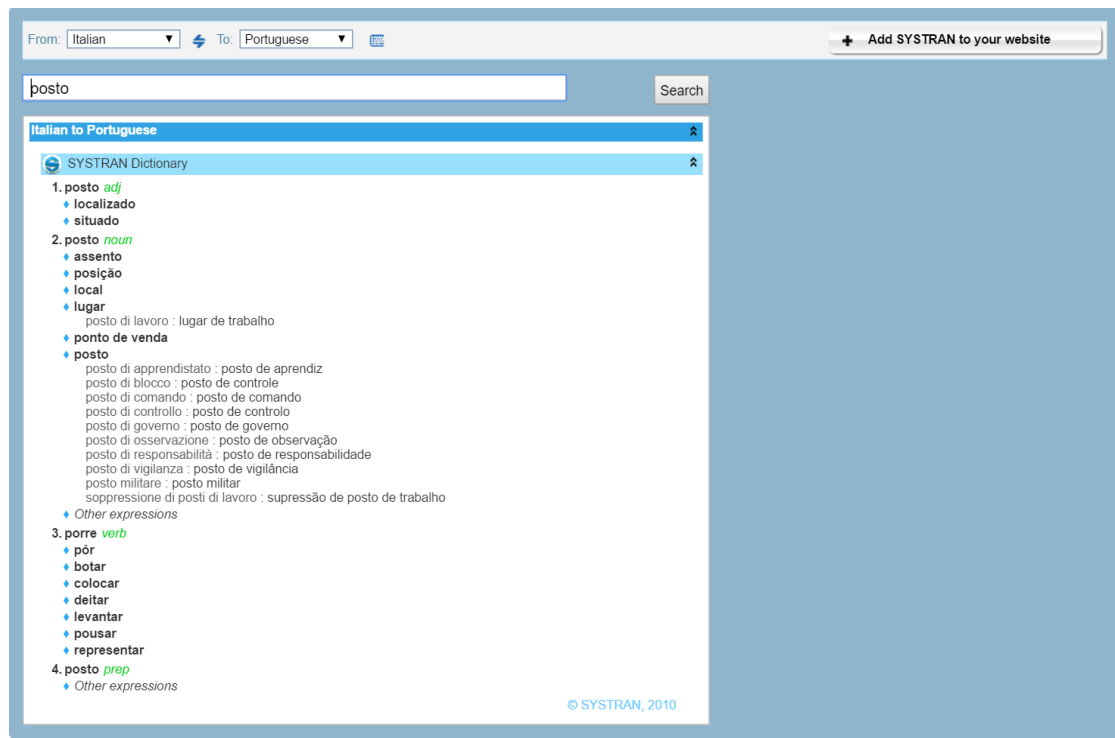


Figura 5. Correspondências de *posto* no dicionário do SystraNet.

Para facilitar a tradução, por conseguinte, é melhor tirar *più* e *esserci*, substituindo este pelo verbo *rimanere* seguido pelo adjetivo *nessun*:

LC:

(6) b. È capace di non rimanere nessun posto. > *É capaz de não permanecer nenhum place.

Porém, embora no dicionário do sistema esteja registada a palavra *posto* com a aceção que assume na frase, o sistema não faz a seleção adequada. Muito provavelmente o sistema utiliza uma língua ponte para a tradução, neste caso o francês (*posto* > *place*). Assim sendo, para obter uma tradução aceitável para português, haverá de utilizar-se o sinónimo mais próximo dessa palavra, ou seja, *luogo*:

LC:

(6) c. È capace di non rimanere nessun luogo. > É capaz de não permanecer nenhum lugar.

LEGGERE > LIGEIRAS. A frase (7) a. é um outro caso de linguagem controlada – a partir da frase (7) – em consequência do problema do imperativo negativo. Além disso,

foi trocado também o pronome objeto indireto *le* pelo pronome objeto indireto de terceira pessoa *gli*, para fazer com que resulte a tradução correta para português:

(7) Non legga quell'articolo! > *Não lê este artigo!

LC:

(7) a. Non gli raccomando di leggere quell'articolo. > *Não lhe recomendo deste ligeiros artigo.

O problema da frase está no verbo *leggere*. De facto, este verbo é homógrafo do adjetivo *leggere* (feminino plural de *leggero*). O dicionário do sistema contem as duas categorias:

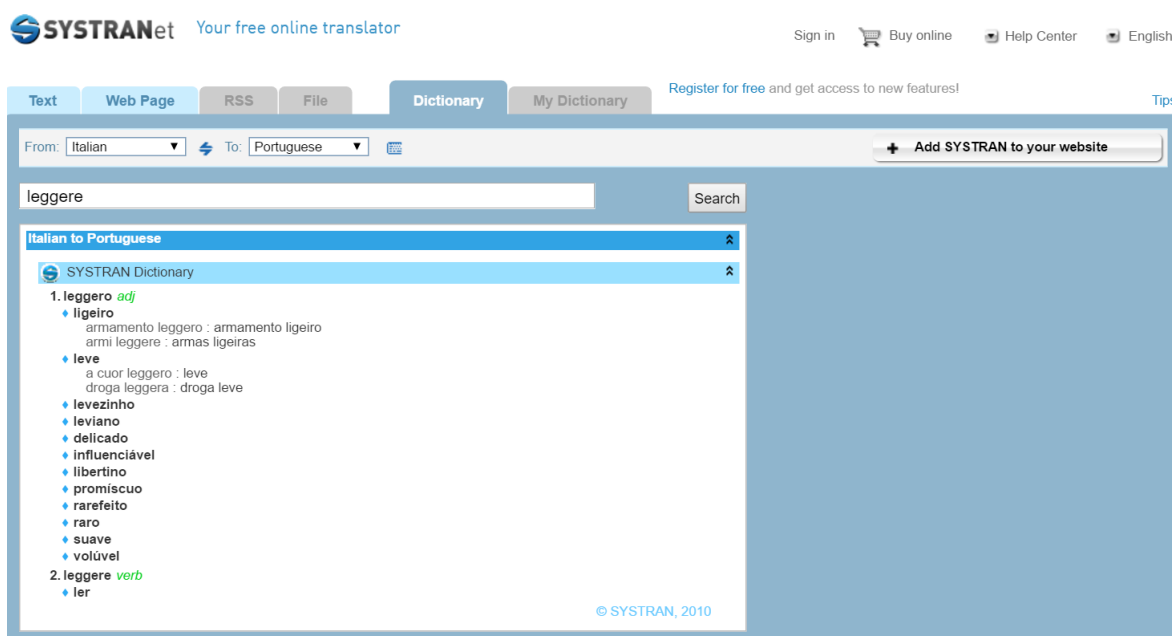


Figura 6. Correspondências de *leggere* no dicionário do SystraNet.

No entanto, o sistema traduz pelo adjetivo. Assim, a solução para controlar a linguagem será substituir o verbo pelo substantivo correspondente, de maneira que seja impossível criar ambiguidade:

LC:

(7) b. Non gli raccomando la lettura di quell'articolo. > Não lhe recomendo a leitura deste artigo.

ACCESA > ACENDIDA. A frase (8) a. foi controlada a partir da (8), sendo que a frase imperativa negativa no *input* não permite um *output* correto:

(8) Non lasciare la luce accesa! > Não deixar a luz acendida!

LC:

(8) a. Non devi lasciare la luce accesa. > *Não debes deixar a luz acendida.

O problema apresenta-se na tradução de *accesa* pelo particípio passado regular de *acender* (“acendida”). De facto, este verbo é um daqueles verbos que apresentam duas formas de particípio (regular: *acendido*; irregular: *aceso*), em que a forma regular é usada nos tempos compostos e a forma irregular como adjetivo ou na voz passiva. Neste caso, a forma adequada é *acesa*. Em contrapartida, em italiano só há uma forma de particípio passado (*accesa*).

O dicionário do SystraNet mostra que *accesso* surge como adjetivo e é traduzido por *aceso* (Figura 7). Assim sendo, a seleção de apenas uma forma do particípio pode ocorrer por duas razões: ou o sistema só gera a forma regular ou, podendo gerar as duas, não tem especificações que lhe permitam seleccionar a adequada.

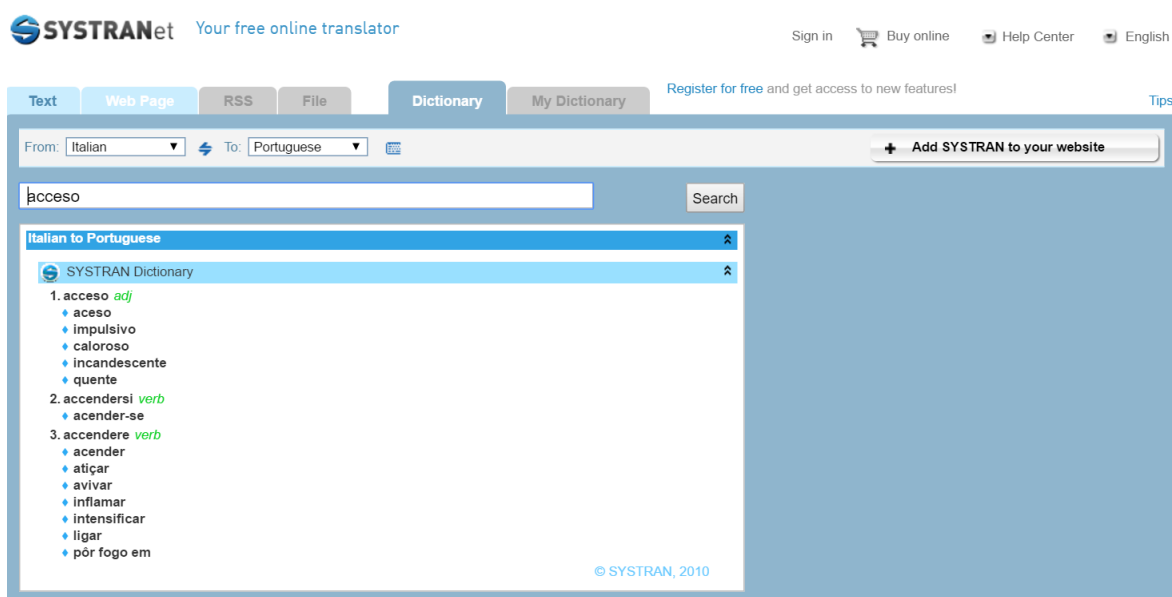


Figura 7. Correspondências de *accesso* no dicionário do SystraNet.

A correspondência está presente e é classificada como adjetivo, mas apesar disso, não há forma de controlar a linguagem na frase (8) a., pois resulta impossível fazer com que o sistema gere a forma irregular do particípio passado.

PRIMA > ANTES, PRIMEIRO, COM ANTECEDÊNCIA.

A frase (9) a. foi controlada substituindo o verbo *dovere* por *avere di no passato prossimo*:

(9) Maria dovette uscire prima per non perdere l'aereo. > *Marie dut sair primeiro para não perder o avião.

LC:

(9) a. Maria ha avuto di uscire prima per non perdere l'aereo. > Marie teve de sair primeiro para não perder o avião.

A frase apresenta uma ambiguidade por causa do advérbio *prima*. É um advérbio com valor temporal que exprime anterioridade em relação a um intervalo de tempo que pode ou não estar expresso na frase. Quando o termo de comparação com o outro intervalo de tempo não há, como neste caso, a tradução para português é *antes*, caso contrário é *primeiro*. O advérbio na frase (9) a. é traduzido por *primeiro*, mas, posto que não há nenhum termo de comparação – a frase encontra-se descontextualizada – e o que se quer expressar aqui é que Maria saiu “com antecedência” em relação à pontualidade do avião, o advérbio apropriado na tradução é *antes*. Obviamente, o resultado da tradução não tem elementos de agramaticalidade, mas o sentido da frase de *output* não corresponde ao da frase de *input*.

No dicionário do SystraNet aparecem as correspondências certas em termos de categorias gramaticais, pois o item em questão surge quer em associação a advérbio quer a preposição, já que *antes* integra a locução preposicional *antes de*:

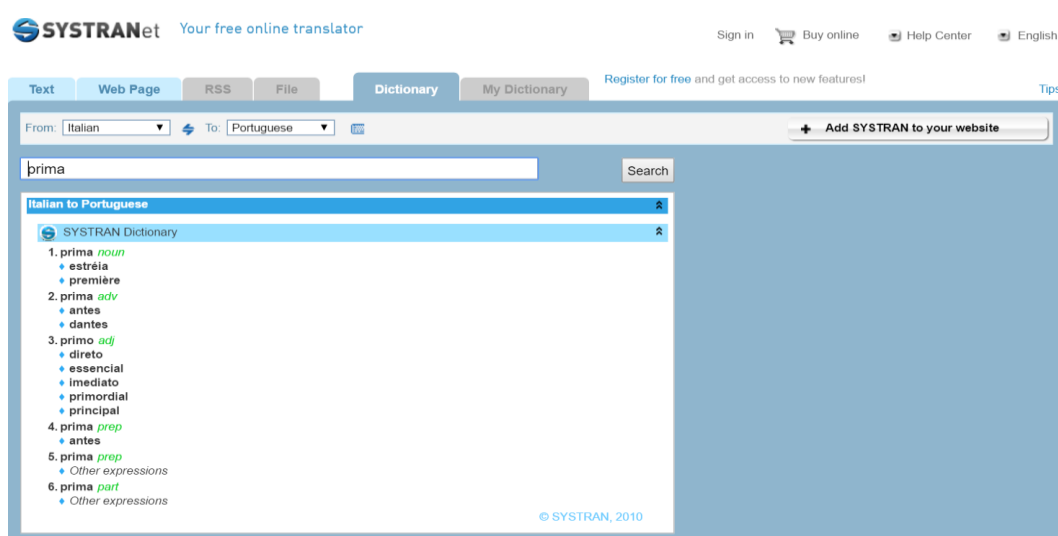


Figura 8. Correspondências de *prima* no dicionário do SystraNet.

O que está em causa é o facto de o sistema escolher a opção *primeiro* com mais frequência, apesar de não estar registada no dicionário:

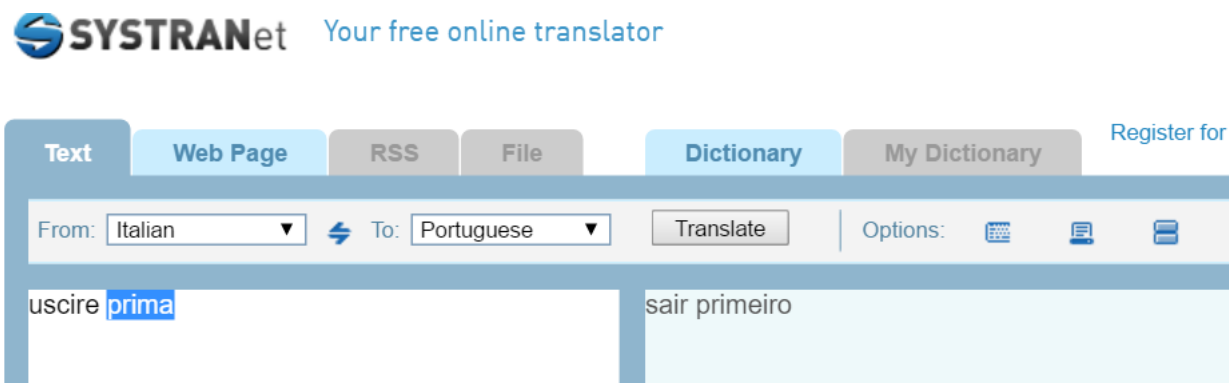


Figura 9. Resultado da tradução de *uscire prima* do SystraNet.

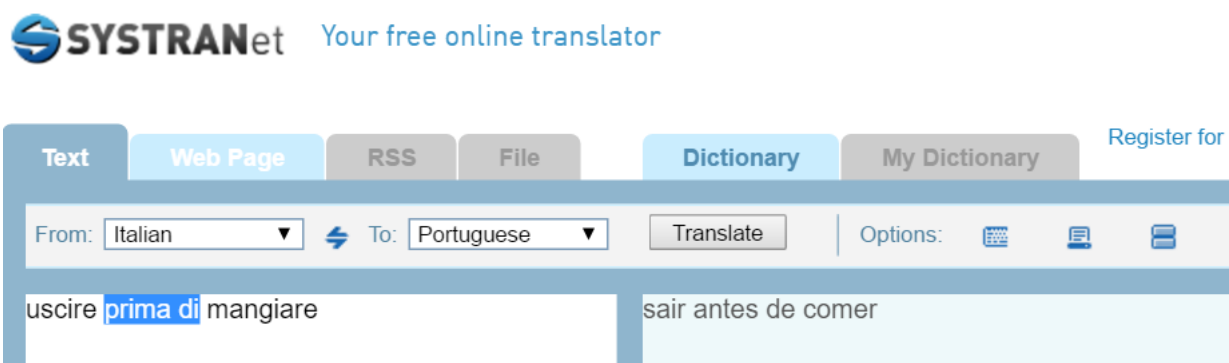


Figura 10. Resultado da tradução de *uscire prima di mangiare* do SystraNet.

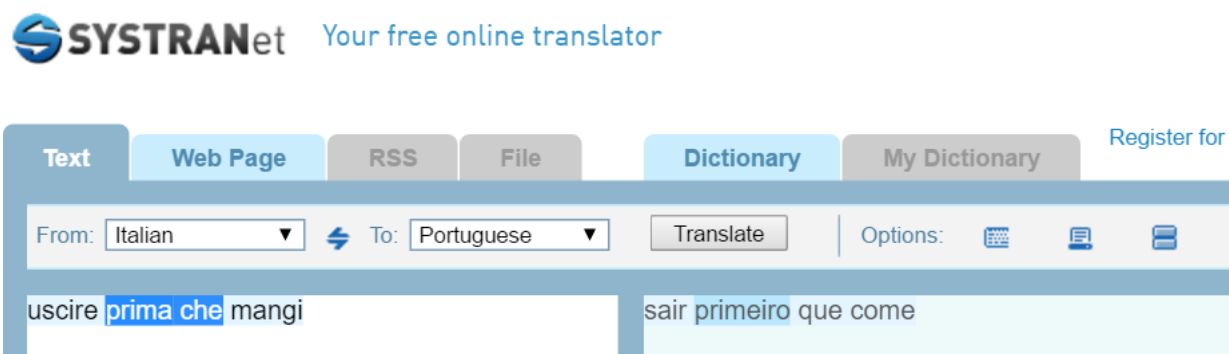


Figura 11. Resultado da tradução de *uscire prima che mangi* do SystraNet.

Como se pode observar nas figuras 9, 10 e 11, o único caso em que o SystraNet seleciona *antes* na tradução é no caso da locução preposicional *prima di*. Nos outros casos escolhe *primeiro*.

Voltando à frase (9) a., dada a impossibilidade de obter uma tradução com o advérbio *antes*, a linguagem poderá ser controlada substituindo *prima* por *con antecedenza*, de maneira a que o resultado da tradução seja correto:

LC:

(9) b. Maria ha avuto di uscire con antecedenza per non perdere l'aereo > Marie teve de sair com antecedência para não perder o avião

Assim sendo, pode estabelecer-se uma regra conforme a ocorrência do tempo de referência. Se este intervalo de tempo não ocorrer de maneira explícita na frase, é preciso substituir *prima* por *con antecedenza*.

PRIMA CHE/PRIMA DI → ANTES QUE/ANTES DE. O item *prima*, como mencionado acima, pode integrar também uma locução conjuncional ou preposicional quando é seguido, respetivamente, por *che* ou *di*. Sendo *prima* indispensável para a formação de frases temporais cuja subordinada tem relação de anterioridade com a principal, envolve problemáticas dificilmente solucionáveis em relação ao modo. De facto, como já mencionado no capítulo 4, há a possibilidade de controlar a linguagem só no caso dos sujeitos de primeira e terceira pessoa do singular ou quando o verbo tem sujeito expletivo nulo, substituindo as frases finitas pelas não finitas correspondentes:

(10) È meglio uscire prima che inizi a piovere. > *É melhor sair primeiro que inicia chover.

LC:

(10) a. È meglio uscire prima di iniziare a piovere. > É melhor sair antes de começar a chover.

(11) Andiamo via prima che lui ritorni. > *Vamos primeiro que retorna.

LC:

(11) a. Andiamo via prima di lui ritornare. > Vamos antes de ele retornar.

Contudo, tomando em consideração as mesmas frases com formas diferentes do verbo da completiva, este tipo de controlo não serve para resolver o problema, sendo que em italiano só existe o infinitivo não flexionado:

(12) Andiamo via prima che loro ritornino.

LC:

(12) a. Andiamo via prima di loro ritornare. > *Vamos antes de eles retornar.

Por conseguinte, é preciso encontrar uma solução alternativa.

Como demonstrado nas figuras 10 e 11, o sistema traduz corretamente só quando *prima* é seguido por *di*, mas mesmo colocando esta preposição entre *prima* e *che* para que o sistema traduza *antes que*, ainda fica o problema da tradução do conjuntivo, seja presente ou passado:

LC:

(12) b. Andiamo via prima di che loro ritornino. > *Vamos antes que retornam.

LC:

(13) Siamo andati via prima di che loro ritornassero. > *Fomos antes que retornavam.

Consequentemente, não há possibilidade de resolver o problema com nos casos em que o verbo da subordinada não tenha sujeito expletivo nulo ou não seja na primeira ou terceira pessoa do singular.

FARE TARDI > CHEGAR TARDE / TARDI > TARDE / VERGOGNARSI > TER VERGONHA DE.

A frase (14) a. foi controlada a partir da (14):

(14) Si vede che ha fatto tardi e si vergognava a dirlo. > *Vê-se que tem feito atrasado e tinha-se vergonha a dizê-lo.

LC:

(14) a. Deve aver fatto tardi e si vergognava a dirlo. > *Deve ter feito atrasado e tinha-se vergonha a dizê-lo.

Nesta frase encontram-se três tipos de problemas: a expressão *fare tardi*, o problema do sistema na tradução de *tardi* e o verbo reflexivo *vergognarsi*.

Fare tardi è uma expressão usada na linguagem informal e não pode ser traduzida literalmente, pois não tem correspondência em português. O sinónimo da expressão *fare tardi* é *arrivare tardi* (“chegar tarde”), pelo que será necessário controlar a linguagem com a substituição da primeira pela segunda:

LC:

(14) b. Deve essere arrivato tardi e si vergognava a dirlo. > *Deve ter chegado atrasado e tinha-se vergonha a dizê-lo.

Contudo, o problema de tradução de *tardi* não se pode resolver, sendo este reconhecido pelo sistema como adjetivo e não como advérbio. De facto, no dicionário, a correspondência *tardi* > *tarde* não está presente, a palavra “tarde” aparece só em forma de adjetivo correspondente de *tardo* (“lento”, “atrasado”):

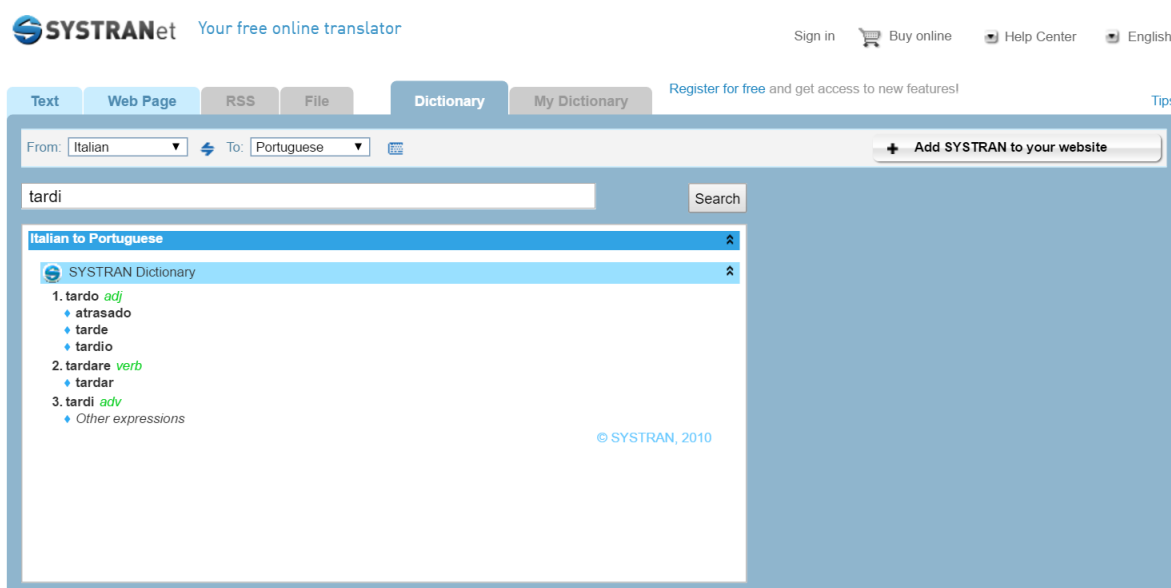


Figura 12. Correspondências de *tardi* no dicionário do SystraNet.

Outro problema que é preciso resolver nesta frase é o verbo *vergognarsi* (“ter vergonha”, “envergonhar-se”), verbo reflexivo que, neste caso, rege a preposição *a*. O SystraNet reconhece que este verbo corresponde a *ter vergonha* mas não seleciona a preposição correta. Em consequência, para controlar a linguagem, será preciso usar o verbo *vergognare* sem o pronome reflexivo e substituir a preposição *a* por *di*, para permitir ao sistema construir corretamente a expressão na língua portuguesa:

LC:

(14) c. Deve essere arrivato tardi e vergognava di dirlo. > ? Deve ter chegado atrasado e tinha vergonha de dizê-lo.

ESSERCI > ESTAR/HAVER.

Esserci compõe-se de *essere* mais *ci*, tendo um significado diferente de *essere*. Assume o significado de “haver”, “existir” e flexiona-se apenas na terceira pessoa (singular e plural), *c’è*, *ci sono*.

Em português este verbo corresponde a *haver* ou *estar*, dependendo dos argumentos que são considerados na frase.

Seguem alguns exemplos:

(15) In frigo ci sono delle uova. → No frigorífico há uns ovos.

(16) A casa c'è Luisa. → Em casa está Luisa.

Em (15), *ci sono* identifica a existência de um objeto num determinado lugar e, por conseguinte, traduz-se por *haver*.

Em contrapartida, em (16) *c'è* não denota existência, mas identifica apenas a presença de qualquer pessoa ou coisa específica (num determinado lugar), da qual se sabe a existência. A partícula *ci*, neste caso, introduz o locativo (“a casa”), pelo que *Luisa* é o sujeito do verbo *essere*. O exemplo (16), por conseguinte, poderá ser invertido da seguinte maneira:

(16) a. Luisa è a casa. → Luisa está em casa.

Toma-se como exemplo a frase (17) a., a qual foi controlada a partir da (17):

(17) Ci dev'essere un bavaglino nella borsa, prendilo! > *Deve-nos ser um babeiro na bolsa, toma!

LC:

(17) a. Ha di esserci un bavaglino nella borsa, prendilo! > *Tem ser um babeiro na bolsa, toma!

Neste caso, o verbo *esserci* denota existência, pelo que deveria ser traduzido por *haver*. Este verbo é o que causa problemas à tradução, sendo que *haver* não consta do dicionário como uma das aceções de *esserci*:

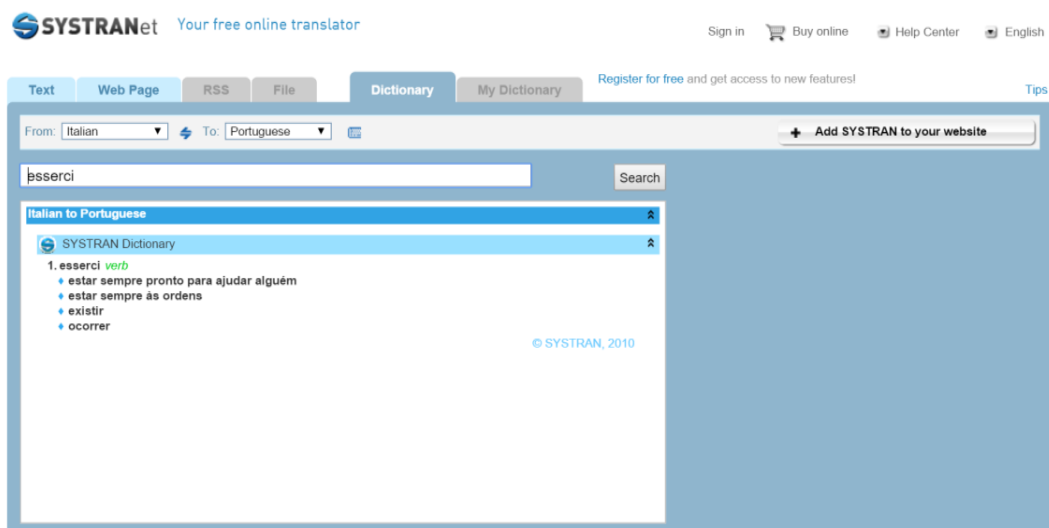


Figura 13. Correspondências de *esserci* no dicionário do SystraNet.

Entre os verbos correspondentes presentes no sistema, encontra-se também *existir*, tendo o verbo *esserci* também tal significado.

Assim sendo, o controlo da linguagem será efetuado com a substituição de *esserci* por *esistere*:

LC:

(17) b. Ha di esistere un bavaglino nella borsa, prendilo! > ? Tem de existir um bafeiro na bolsa, toma!

No entanto, para controlar a ocorrência do verbo *esserci* não se pode estabelecer nenhuma regra específica. Este verbo pode aparecer em diversos tipos de contexto e, em consequência, as suas correspondências em português também podem ser diferentes. Por exemplo, tomando em consideração as frases (15) e (16), dever-se-ia efetuar um controlo diferente, pois o verbo da primeira frase exprime existência e o verbo da segunda introduz um locativo:

(15) In frigo ci sono delle uova. > Em refrigerador há ovos.

LC:

(16) a. Nel⁷¹ frigo ci sono delle uova. > No refrigerador há ovos.

Neste caso, o sistema reconhece *ci sono* como denotando existência, contrariamente ao que acontece com *esserci* na forma infinitiva.

⁷¹ Substituição da preposição simples *in* pela contraída *nel*, por causa do português “no refrigerador”

No caso de (16), a questão é diferente:

(16) A casa c'è Luisa. > *À casa ele lá Luisa.

Como se pode observar, aqui ocorre o mesmo problema da frase (17), ou seja, a falta de reconhecimento do verbo *esserci*. Assim sendo, para controlar a linguagem, será preciso alterar a ordem da frase:

LC:

(16) b. Luisa é dentro⁷² casa. > Luisa está em casa.

PRENDERE > ENCONTRAR. Na frase (17) b. foi levantado também um problema pragmático que tem a ver com a expressão imperativa *prendilo!* (“pega nele!”). A tradução dada pelo sistema (“toma!”) não é adequada. A frase, em italiano, pressupõe o carácter factual do evento descrito na frase anterior, o que a tradução (literal) não deixa ver. Há, por conseguinte, que controlar esta frase por forma a obter um resultado que dê conta da referida pressuposição.

Em consequência, a frase em linguagem controlada será a seguinte:

LC:

(17) c. Ha di esistere un bavaglino nella borsa, lui lo trova! > Tem de existir um bafeiro na bolsa, encontra-o!

Neste caso, nenhuma regra pode ser estabelecida, sendo que a unidade lexical traduzida pelo sistema (*tomar*) não permite o efeito pragmático por causa do problema – já analisado – da polissemia do verbo *prendere*. Todavia, a substituição feita na frase controlada (16) c. não permite a criação de nenhuma regra por ter sido efetuada em relação à sua função comunicativa.

ESSERE IN GRADO > ESTAR EM CONDIÇÕES. A frase (18) a. foi controlada a partir da (18) e foi substituída pela forma finita por não existir em português a correspondente não finita:

(18) Penso di non essere in grado di terminare il lavoro. > *Penso não ser em condições da terminar trabalho.

LC:

⁷² Controlo do fenómeno relativo aos complementos locativos, tratado na secção 7.2.

(18) a. Penso che io non sono in grado di terminare il lavoro. > Penso que não sou em condições de terminar o trabalho.

Essere in grado é uma expressão da língua italiana que em português pode ser traduzida por “estar em condições” ou, mais simplesmente, “ser capaz” entendido como habilidade. O sistema não está longe da tradução certa, pois considera literalmente o verbo *essere* como *ser* em vez de *estar*, o qual parece ser o único obstáculo para a tradução. Assim sendo, para controlar a linguagem é preciso substituir esta expressão por *essere capace*, no seu verdadeiro sentido de habilidade:

LC:

(18) b. Penso che io non sono capace di terminare il lavoro. > Penso que não sou capaz de terminar o trabalho.

DEMONSTRATIVOS. A frase (19), já tratada precedentemente na mesma secção para a questão da ambiguidade do verbo *leggere*, apresenta também problemas de correspondência relativamente aos adjetivos demonstrativos.

(19) Non gli raccomando la lettura di quell'articolo. > *Não o recomendo a leitura deste artigo.

Como se pode observar, o sistema não reconhece a correspondência certa: *quello* corresponde a *aquele* e *questo* corresponde a *este*. Em italiano não existe a forma de demonstrativo que em português é expressa por *esse*, pelo que pode ser esta a diferença que causa problemas ao sistema.

Para controlar a linguagem desta frase, pode tentar-se substituir o adjetivo demonstrativo *quell'* pela sua forma inteira *quello*:

LC:

(19) a. Non gli raccomando la lettura di quello articolo. → *Não lhe recomendo a leitura de aquilo articulo.

Nesta frase encontramos dois elementos que tornam a frase agramatical: *quello* é traduzido por *aquilo* e *articolo* por *articulo*. *Aquilo* é o pronome demonstrativo, pelo que não pode anteceder um nome, e *articulo* é uma tradução errada da palavra *artigo*.

No entanto, no dicionário do SystraNet estão as correspondências certas:

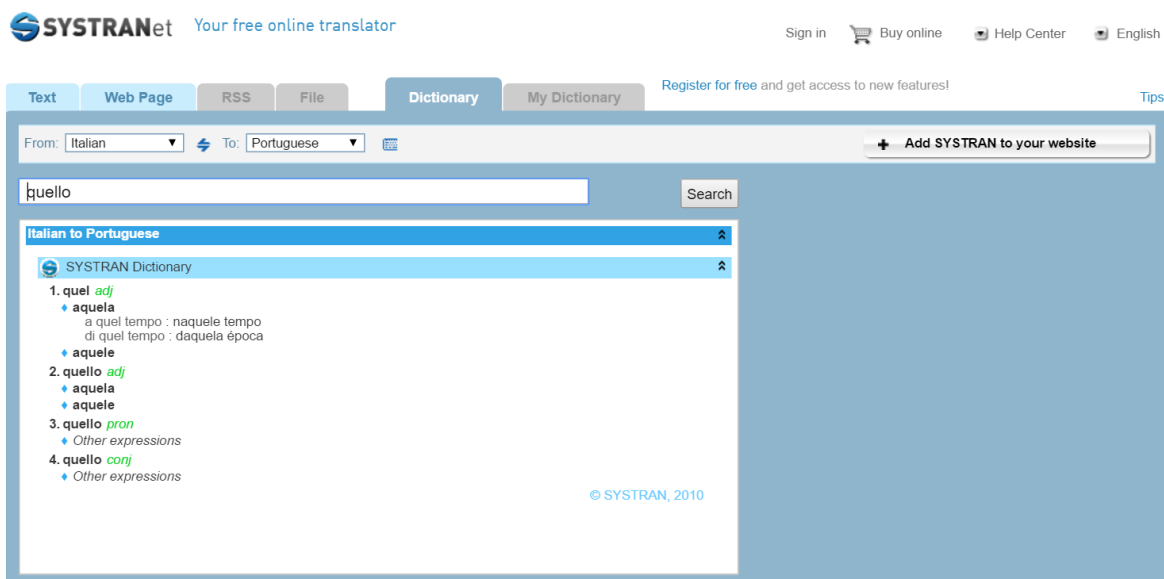


Figura 14. Correspondências de *quello* no dicionário do SystraNet.

Como se pode verificar, as correspondências dos adjetivos *quel* e *quello* são corretas (*aquele* e *aquela*). Apesar disso, não há maneira de controlar a linguagem, sendo que o sistema traduzirá sempre por *este* ou *aquilo* e nunca por *aquele*.

7.2 ASPETOS SINTÁTICOS

A CASA SUA > EM CASA. A partir da frase (20), foi controlada a linguagem na (20) a.:

(20) Può darsi che Andrea sia a casa sua. > *Pode dar-se que Andrea está à casa o seu.

LC:

(20) a. Probabilmente Andrea è a casa sua. > * Provavelmente Andrea está à casa o seu.

O problema desta frase é o complemento locativo *a casa sua*, o qual tem uma estrutura diferente em português. Com efeito, a língua portuguesa diferencia o complemento locativo de situação do direcional através das preposições que o antecedem.

Em italiano, esta diferença não é determinada pelas preposições, mas apenas pelos verbos. As preposições são determinadas pelo lugar e o seu uso depende de vários fatores. Veja-se a diferença:

(21) Sono a casa. → Estou em casa. (Complemento locativo de situação)

(22) Vado a casa. → Vou a casa. (Complemento locativo direcional)

Por conseguinte, para controlar a frase (20), será preciso usar a preposição *in* (“em”).

Dado que em (20) também está o adjetivo possessivo (*sua*), é preciso pô-lo antes do substantivo para respeitar a estrutura sintática do português:

LC:

(20) b. Possivelmente Andrea è in sua casa. > Possivelmente Andrea está na sua casa.

O resultado da tradução, embora não seja uma frase agramatical, precisa de um controlo adicional para a obtenção de *em casa* no *output*:

LC:

(20) c. Possivelmente Andrea è in casa. > *Possivelmente Andrea está ø casa.

Neste caso, ao tirar o adjetivo possessivo, o sistema não traduz a preposição *em*. Este fenómeno ocorre também quando a preposição é seguida por um nome próprio de cidade:

(23) Mario è a Roma. > *Mario está à Roma.

LC:

(23) a. Mario è in Roma. > *Mario está ø Roma.

Por conseguinte, a única maneira de controlar a linguagem é, uma vez mais, substituir a preposição, mas desta vez por um sinónimo de *in*, *dentro*. Esta preposição dita “acidental” (*preposizione impropria* em italiano) pode substituir a preposição dita “essencial” *in*, por ter a mesma importância seja no aspeto morfológico, seja no aspecto sintático (Zuccarello e Belizário, 2009). O SystraNet reconhece tal preposição e executa uma tradução correta:

LC:

(23) b. Mario è dentro Roma. > Mario está em Roma.

Desta maneira, testa-se o desempenho do sistema com a substituição de *in* por *dentro*:

(20) Possivelmente Andrea è a casa.

LC:

(20) d. Possivelmente Andrea è dentro casa. > Possivelmente Andrea está em casa.

(24) Giovanni è al cinema.

LC:

(24) a. Giovanni è dentro il cinema. > Giovanni está no cinema.

(25) Luisa è a teatro.

LC:

(25) a. Luisa è dentro il teatro. > Luisa está no teatro.

Assim sendo, pode concluir-se que no caso de locativos de situação introduzidos pela preposição *a*, é preciso substituir tal preposição (*a*) por *dentro*.

ANDARE A CASA > IR A CASA. Toma-se em consideração, novamente⁷³, a frase

(26):

(26) Penso che Giovanni non ritorna a casa oggi. > ? Penso que Giovanni não retorna à casa hoje.

O resultado de tradução de (26) tem um pequeno detalhe que obstacula a aceitabilidade da frase, a preposição *à* (“à casa”). O sistema parece associar diretamente a preposição italiana *a* à contração da preposição *a* com o determinante *a* (*à*) em português, sem nenhuma razão aparente. Sendo a expressão portuguesa “a casa” mais correta do que “à casa”, é preciso controlar a linguagem com o artigo sem preposição, “la casa”, para que na tradução ocorra só a preposição. Este controlo funciona porque *à* é o resultado da contração da preposição *a* mais o artigo definido *a* e os equivalentes em italiano são *a* (*preposizione semplice*) mais o artigo definido *la*. Dado que com *a* o sistema não traduz corretamente, a única solução será substituir *a* por *la*:

LC:

(26) a. Penso che Giovanni non ritorna la casa oggi. > Penso que Giovanni não retorna a casa hoje.

IN > A. A frase (27) tem um problema sintático em relação ao complemento locativo:

(27) Mio padre arrivò in Italia nel 1956. > *O meu pai chegou na Itália em 1956.

⁷³ Frase já considerada na secção 7.1 pelo problema lexical levantado pelo verbo *ritornare*.

Como referido precedentemente, em italiano não é através o uso das preposições que se determina se o complemento locativo é direcional ou não. Em contrapartida, em português a preposição *em* ocorre no complemento locativo de situação e *a* no complemento locativo direcional. Neste caso, o verbo *arrivare* (*chegar*) denota movimento e, em consequência, requer um locativo direcional.

Assim sendo, para controlar a linguagem, é preciso substituir a preposição *in* por *a*, quando o verbo denota movimento.

No entanto, como constatado anteriormente, o SystraNet traduz *a* pela contração da preposição *a* com o determinante (*à*), em todos os casos:

LC:

(27) a. Mio padre arrivò a Italia nel 1956. > O meu pai chegou à Itália em 1956.

Há aqui uma questão controversa em relação ao uso do artigo definido antes dos países: na língua portuguesa há países que são precedidos pelo artigo definido e outros não. No caso de Itália, assim como Espanha, França e mais alguns, podem ser ou não precedidos pelo artigo, pelo que em (27) a. o resultado de tradução fica correto. Em contrapartida, países como Portugal, Marrocos, Moçambique e mais alguns, não aceitam o artigo definido. Em relação às cidades, o discurso é oposto: geralmente não é aceite o artigo definido, mas há alguns casos em que é obrigatório (por exemplo, a Porto). Assim sendo, no caso da frase (27) a., o controlo da linguagem pode ser efetuado para melhorar o *output*, mas tal controlo não se pode tornar regra pelo facto de não valer para todos os casos. Uma solução possível poderia ser aplicar a regra acima mencionada e então substituir *a* por *la* nos casos que rejeitam o artigo, para que fique a preposição sem contração com o determinante:

LC:

(27) b. Mio padre arrivò la Italia nel 1956. > O meu pai chegou a Itália em 1956.

O problema deste tipo de controlo é que nem sempre funciona com outros países ou cidades:

(28) Sono andato a Lisbona. > *Fui à Lisboa.

LC:

(28) a. Sono andato la Lisbona. > *Fui ø Lisboa.

(29) Sono andato in Portogallo. > Fui à Portugal.

LC:

(29) a. Sono andato a Portogallo. > *Fui à Portugal.

LC:

(29) b. Sono andato la Portogallo. > *Fui ø Portugal.

Como se pode notar nas frases controladas (28) a. e (29) b., o sistema omite o artigo/preposição. Em (28), (29) e (29) a. demonstrou-se que o determinante não se adequa ao nome que tem a seguir. Provavelmente o analisador morfológico do sistema não tem género para nomes próprios e – como já observado – deve ter também uma tendência em traduzir a preposição *a* por *à*, mesmo quando não é de género feminino.

Por estas razões, não é possível resolver o problema com a linguagem controlada. Poderiam, contudo, estabelecer-se algumas restrições para apenas algumas aplicações, mas deveriam sempre requerer um controlo humano do *output*, sendo que o uso do locativo direcional, como demonstrado, vai ser sempre problemático na tradução automática.

No entanto, neste tipo de trabalho não cabe a análise e a criação de regras que exigem pós-edição.

IN TEMPO > A TEMPO / GIOCARE A TENNIS > JOGAR TÊNIS.

Toma-se em

consideração a frase (30):

(30) Se arriviamo in tempo andremo a giocare a tennis insieme. > *Se chegarmos em tempos iremos jogar ao ténis juntos.

Esta frase é traduzida corretamente em relação ao modo. Com efeito, o sistema traduz o verbo no presente do indicativo depois do *se* pelo verbo no futuro do conjuntivo em português. Todavia, na tradução da frase encontramos outros problemas. Primeiro, a locução adverbial temporal *in tempo* em português deveria ser traduzida por *a tempo*, mas, não reconhecendo esta expressão como tal, traduz por *em tempos*. Além disso, apresenta-se outro problema devido à preposição *a* na expressão *giocare a tennis*: esta preposição não deveria estar presente na frase em português porque o verbo jogar é transitivo direto, pois seleciona um complemento nominal, sem preposição.

Para controlar *in tempo*, é preciso substituir a preposição *in* por *a*; para controlar *giocare a tennis* é preciso apenas tirar a preposição *a*:

LC:

(30) a. Se arriviamo a tempo andremo a giocare \emptyset tennis insieme. > Se chegarmos a tempo iremos jogar ténis juntos.

VERBO NA SEGUNDA PESSOA DO PLURAL > VERBO NA TERCEIRA PESSOA DO PLURAL.

Consideram-se as frases com o verbo na segunda pessoa do plural que causaram problemas ao sistema:

(31) Chiama appena siete pronti. > *Chama logo que estiver pronti.

(32) Andate a casa! > *Aleia à casa!

Em (31), como se pode verificar, o sistema traduz o verbo no tempo e no modo corretos (futuro do conjuntivo), mas na pessoa errada (terceira do singular em vez da terceira pessoa do plural). Em (32) o verbo está no modo imperativo, o que constitui um problema adicional, como demonstrado pelo resultado de tradução. O sistema não reconhece a correspondência *voi*>*vós*, provavelmente porque a segunda pessoa do plural (“vós”) tem um uso limitado em português⁷⁴.

Veja-se uma frase simples com o verbo na segunda pessoa do plural no presente do indicativo:

(33) Voi mangiate troppo pesce. > Come demasiado peixe.

Como se pode observar, também em (33) o sistema traduz pela terceira pessoa do singular. Sendo que em português a forma correspondente da segunda pessoa do plural (*voi*) é a terceira pessoa do plural (*vocês*), o controlo pode efetuar-se com a substituição do verbo na segunda pessoa do plural pelo verbo na terceira pessoa do plural, especificando o sujeito do verbo (*loro*):

(31) a. Chiama appena loro sono pronti. > Chama logo que estiverem pronti.

(32) a. Loro vanno a casa! > Vão a casa!

(33) a. Loro mangiano troppo pesce. > Comem demasiado peixe.

⁷⁴ A segunda pessoa do plural “vós” usa-se em textos muito formais e em certos dialetos da região Norte de Portugal.

7.3 CONCLUSÃO

Em conclusão, os fenómenos que criam problemas à tradução automática decorrem das especificidades da língua. Na maioria dos casos analisados, as diferenças linguísticas entre italiano e português obstacularam a obtenção de resultados de tradução satisfatórios. Foram levantados casos em que ambiguidades lexicais causaram uma má tradução (como no caso de *prima* ou *leggere*) e também casos de tradução incorreta causada apenas pelo mau desempenho do sistema.

Para resolver estes problemas, foi aplicado um controlo onde possível e também foram estabelecidas algumas regras específicas para permitir o melhoramento dos resultados na língua de *output*.

No entanto, encontram-se muitas incongruências entre o dicionário do SystraNet e a tradução de um texto no sistema. De facto, houve casos em que não se pôde resolver o problema de tradução através da linguagem controlada, pois resultaram correspondências não adequadas ao contexto (por exemplo no caso de *prendere*). Assim sendo, nos casos em que não foi possível controlar a linguagem para obter um *output* adequado e, conseqüentemente, nem a criação de determinadas regras, foram levantados os problemas demonstrando também o mau funcionamento do sistema.

Contudo, é preciso salientar novamente que as regras criadas não são exaustivas, pois são apenas casos encontrados na análise de modo, modalidade e aspeto.

8. CONCLUSÕES E CONSIDERAÇÕES FINAIS

O presente trabalho teve como principal objetivo o estabelecimento de regras de linguagem controlada para melhorar os resultados da tradução automática de italiano para português, tendo-se feito uso, para o efeito, do sistema SystraNet. Tais regras foram criadas a partir da análise das variações linguísticas dos três grandes tópicos tomados em consideração neste trabalho, nomeadamente modo, modalidade e aspeto. No entanto, foram analisados também outros fenómenos que criam problemas à tradução automática e que foram encontrados ao longo da análise.

Os objetivos justificam-se com a necessidade de obter um italiano controlado para melhorar, tanto quanto possível, a qualidade dos resultados da tradução automática, num sistema com uma forte componente orientada para o conhecimento linguístico. Além disso, esta escolha surge do facto de o par linguístico italiano-português ser ainda pouco explorado na área das linguagens controladas para a tradução automática.

Desta maneira, no capítulo 2 foram apresentados os modelos de tradução automática, com um particular aprofundamento dos sistemas híbridos.

No capítulo a seguir apresentaram-se as linguagens controladas e as suas classificações e aplicações, em particular no âmbito da tradução automática.

Nos três capítulos a seguir foram tratadas as variações das duas línguas no âmbito de modo, modalidade e aspeto, o núcleo da análise deste trabalho. Ao longo destes capítulos, foi possível notar como as diferenças entre as duas línguas podem ser ou não relevantes na tradução automática. Sendo o SystraNet um sistema baseado principalmente em regras, em alguns casos reconhece essas diferenças e o resultado da tradução é correto. Em outros casos, em contrapartida, as regras do sistema não chegam para obter uma tradução correta, pelo que é preciso efetuar o controlo da linguagem. Também se verificaram casos em que, embora não haja diferenças entre as duas línguas, o sistema traduz de maneira inadequada, pelo que tentou adotar-se métodos de controlo diferentes. Em tais casos, nem sempre foi possível controlar a linguagem e estabelecer regras, pelo que deverão ser tratados na fase de pós-edição (não tratada no presente trabalho).

Enfim, no sétimo capítulo, respeitante a outros fenómenos a requerer controlo, foi demonstrado que nem sempre se podem resolver os problemas da tradução automática

através de regras, pois nem todas as limitações dos sistemas são ultrapassáveis por esta via.

As regras que foram criadas ao longo deste trabalho constituem-se como restrições, apresentadas de forma declarativa, por forma a permitir ao utilizador o uso apropriado das mesmas e a obtenção de resultados relativamente satisfatórios. O conjunto de regras é apresentado no anexo, que está dividido em secções, em que, por seu turno, são apresentadas as regras correspondentes aos diferentes fenómenos linguísticos.

Cabe, mais uma vez, salientar os limites deste trabalho em termos do objeto de estudo, necessariamente confinado a alguns tópicos linguísticos, dado o facto de o mesmo se desenvolver no quadro de um trabalho final de mestrado. A análise linguística para estabelecer este conjunto de regras baseia-se exclusivamente no modo, na modalidade, no aspeto e em alguns fenómenos extra que impediram um bom desempenho do sistema.

Espera-se que este trabalho represente uma contribuição para o incremento da cobertura do italiano controlado para a tradução automática, que nesta fase é ainda de fraca amplitude.

Pelas diversas razões apontadas, considera-se a possibilidade de, no seguimento do presente trabalho, se levar por diante a extensão do italiano controlado a outros fenómenos, noutros contextos de investigação.

ANEXO

REGRAS PARA A LINGUAGEM CONTROLADA ITALIANO – PORTUGUÊS

MODO

1. FRASES IMPERATIVAS

REGRA 1.1

Nas frases imperativas afirmativas, substituir o verbo no modo imperativo nas segundas pessoas (do singular e do plural) pelo verbo imperativo nas terceiras pessoas (do singular e do plural).

Realizar o sujeito do verbo.

Dormi bene!

Lui dorme bene!

Fai buon viaggio!

Lui fa un buon viaggio!

Andate a casa!

Loro vanno la casa!

REGRA 1.2

Substituir as frases imperativas negativas por estruturas não imperativas semanticamente equivalentes.

Non lasciare la luce accesa!

Non devi lasciare la luce accesa.

Non mi portare a casa!

La mia richiesta è di non portarmi la casa.

Non legga quell'articolo!

Non gli raccomando la lettura di quell'articolo.

REGRA 1.3

Nas frases imperativas afirmativas, omitir o pronome clítico ou substituí-lo pela correspondente forma precedida de preposição (a).

Portami un caffè!

Lui porta (a me) un caffè!

Fammi vedere!

Lui fa (a me) vedere!

2. CONJUNTIVO EM FRASES SIMPLES

REGRA 2.1

Não usar o presente do conjuntivo com função dubitativa em frases simples, substituí-lo pelo futuro simples.

Che sia una bugia?

Sarà una bugia?

Che sia Giacomo alla porta?

Sarà Giacomo alla porta?

REGRA 2.2

Colocar o advérbio forse depois do verbo no modo indicativo, precedido pela vírgula.

Forse Giulia è in ufficio.

Giulia è nell'ufficio, forse.

Forse tu non conosci mia sorella.

Tu non conosci mia sorella, forse.

Forse Giovanni e Luca non hanno cenato.

Giovanni e Luca non hanno cenato, forse.

3. FRASES COMPLETIVAS

REGRA 3.1

Nas frases completivas finitas com verbos de opinião, substituir o verbo trovare por pensare, conjugado no modo indicativo e no tempo adequado.

Trovo che Maria stia meglio.

Penso che Maria sta meglio.

REGRA 3.2

Nas frases completivas finitas com verbos epistêmicos e com negação na frase principal, colocar a negação na frase completiva.

Non penso che David pratichi abbastanza sport.

Penso che David non pratichi abbastanza sport.

REGRA 3.3

Nas frases completivas não finitas introduzidas por *di*, substituir a forma não finita com a estrutura *di* + verbo no infinitivo pela forma finita com a estrutura *che* + verbo no indicativo, conjugado no tempo adequado.

Giulia crede di stare bene.
Giulia pensa che lei sta bene.

4. FRASES TEMPORAIS

REGRA 4.1

Nas frases temporais que exprimem anterioridade, substituir *quando* por *appena*, seguido pelo verbo no presente ou no futuro do indicativo.

Quando finisci di lavorare, vienimi a prendere.
Appena finisci di lavorare, lui viene a prendermi.

Chiama quando sei pronto.
Chiama appena tu sei pronto.

Chiama quando siete pronti.
Chiama appena loro sono pronti.

REGRA 4.2

Nas frases temporais que exprimem posterioridade, substituir a forma finita *prima che* + verbo no conjuntivo pela forma não finita *prima di* + verbo no infinitivo, só quando o verbo tem sujeito expletivo nulo, quando está na primeira ou na terceira pessoa do singular.

È meglio uscire prima che inizi a piovere.
È meglio uscire prima di iniziare a piovere.

Andiamo via prima che lui torni.
Andiamo via prima di lui ritornare.

5. FRASES CONDICIONAIS

REGRA 5

Nas frases condicionais factuais, eliminar se, conjugar o verbo da proposição condicional no gerúndio presente e especificar o sujeito na proposição principal.

Se mangi molti dolci, ingrassi.

Mangiando molti dolci, tu diventi grasso.

Se l'acqua arriva alla temperatura di 100 gradi, entra in ebollizione.

Arrivando alla temperatura di 100 gradi, l'acqua entra in ebollizione.

6. FRASES CONCESSIVAS

REGRA 6.1

Nas frases concessivas factuais, substituir anche se + verbo no presente do indicativo por benché + verbo no presente do conjuntivo.

Anche se ho ragione, non voglio insistere.

Benché io abbia ragione, non voglio insistere.

Anche se mangi molti dolci, non diventi grasso.

Benché tu mangi molti dolci, non sei grasso.

REGRA 6.2

Nas frases concessivas, tanto factuais como não factuais, substituir sebbene por benché, mantendo o verbo no modo conjuntivo.

Sebbene possa piovere, usciremo lo stesso.

Benché possa piovere, usciremo nella stessa.

Sebbene studi molto, non ha ancora passato l'esame.

Benché studi molto, ancora non ha passato l'esame.

MODALIDADE

7. VERBO *DOVERE*

REGRA 7.1

Substituir o verbo *dovere* quando denota necessidade/obrigatoriedade ou elevada probabilidade por *avere di*.

Franco deve uscire per comprare le sigarette.

Franco ha di uscire per comprare le sigarette.

Questo quadro deve essere di Picasso, non ho dubbi.

Questo quadro ha di essere di Picasso, io non ho dubbi.

REGRA 7.2

Substituir o verbo *dovere no passato remoto e passato prossimo* por *avere di no passato prossimo*.

Maria dovette uscire prima per non perdere l'aereo.

Maria ha avuto di uscire con antecedenza per non perdere l'aereo.

Ho dovuto insistere ma alla fine l'ho convinto.

Ho avuto di insistere ma alla fine l'ho convinto.

8. EXPRESSIONES MODAIS EPISTEMICAS

REGRA 8.1

Substituir as expressões *può essere che* e *può darsi che* + verbo no conjuntivo pelo advérbio *possibilmente* + verbo no indicativo.

Ma forse un'aspirina? Si può darsi tu abbia ragione.

Ma forse un'aspirina? Sì, possibilmente tu hai ragione.

Può essere che Andrea sia a casa sua.

Possibilmente Andrea è dentro casa.

REGRA 8.2

Substituir a expressão *si vede che* + verbo no indicativo por *dovere* + verbo no infinitivo.

Si vede che aveva molta fame.

Deve aver avuto molta fame.

Si vede che ha fatto tardi e si vergognava a dirlo.

Deve essere arrivato tardi e vergognava di dirlo.

REGRA 8.3

Substituir a expressão mi sa che + verbo no indicativo por penso che + verbo no indicativo.

Mi sa che Giovanni non torna a casa oggi.

Penso che Giovanni non ritorna la casa oggi.

Mi sa che mio fratello va al cinema stasera.

Penso che mio fratello va al cinema stasera.

REGRA 8.4

Substituir a expressão capace che + indicativo por ser capaz de + infinitivo, com ser na forma apropriada ao contexto.

Capace che arrivi domani.

Lui è capace di arrivare domani.

Capace che non c'era più posto.

È capace di non restare nessun luogo.

ASPETO

9. TEMPO *PASSATO REMOTO* – EXPRESSÃO DE PERFETIVIDADE

REGRA 9

Substituir o tempo *passato remoto* pelo *passado prossimo*, para expressar o aspeto perfetivo.

L'anno scorso viaggiavi per l'America.

L'anno scorso ho viaggiato per l'America.

Manzoni nacque nel 1785.

Manzoni è nato nel 1785.

10. TEMPO *PASSATO PROSSIMO* – EXPRESSÃO DE ITERATIVIDADE

REGRA 10

Substituir o tempo *passato prossimo* por *tenere no tempo presente + participio pasado* para expressar o aspeto imperfetivo de um evento iterativo.

Ultimamente ho mangiato troppo.

Ultimamente tengo mangiato troppo.

Negli ultimi giorni Maria é arrivata sempre puntuale.

Negli ultimi giorni Maria tiene arrivato sempre puntuale.

11. PERÍFRASES PROGRESSIVAS

REGRA 11

Substituir as estruturas perifrásticas progressivas *stare + verbo no gerúndio* e *stare a + verbo no infinitivo* pela estrutura *stare per + verbo no infinitivo* para expressar o aspeto progressivo nas frases afirmativas.

Marco adesso sta mangiando.

Marco adesso sta per mangiare.

Quando Elena è entrata in casa, Maria stava mangiando una mela.

Quando Elena è entrata in casa, Maria stava per mangiare una mela.

Paolo è stato a parlare con Luca per tutta la sera.

Paolo stette per parlare con Luca per tutta la sera.

OUTROS FENÓMENOS A REQUERER CONTROLO

12. VERBO NA SEGUNDA PESSOA DO PLURAL

REGRA 12

Substituir o verbo na segunda pessoa do plural pelo verbo na terceira pessoa do plural, com realização do sujeito.

Chiama appena siete pronti.

Chiama appena loro sono pronti.

Voi mangiate troppo pesce.

Loro mangiano troppo pesce.

13. PREPOSIÇÃO “A”

REGRA 13.1

Substituir *a* por *dentro*, quando a preposição introduz um complemento locativo.

Possibilmente Andrea è a casa sua.

Possibilmente Andrea è dentro casa.

Mario è a Roma.

Mario è dentro Roma.

REGRA 13.2

Substituir *a* por *la* quando introduz um complemento locativo direcional sendo o lugar de destino expresso por *casa*.

Penso che Giovanni non torna a casa oggi.

Penso che Giovanni non ritorna la casa oggi.

REGRA 13.3

Eliminar *a* quando é precedido pelo verbo *giocare* e precede a designação do jogo.

Se arriviamo in tempo andremo a giocare a tennis insieme.

Se arriviamo a tempo andremo a giocare \emptyset tennis insieme.

14. NOMES

REGRA 14

Substantivo *posto*: substituir por *luogo*.

È capace di non esserci più posto.
È capace di non rimanere nessun luogo.

15. VERBOS

REGRA 15.1

Verbo *ingrassare*: substituir pelo complexo verbal *diventare grasso* em frases afirmativas.

Se mangi molti dolci, ingrassi.
Mangiando molti dolci, diventi grasso.

REGRA 15.2

Verbo *leggere* no infinitivo: substituir pelo substantivo *lettura*.

Non le raccomando di leggere quell'articolo.
Non gli raccomando la lettura di quell'articolo.

REGRA 15.3

Verbo *tornare*: substituir por *ritornare*.

Penso che Giovanni non torna a casa oggi.
Penso che Giovanni non ritorna la casa oggi.

REGRA 15.4

Verbo reflexivo *vergognarsi*: substituir por *vergognare* + preposição *di* conjugado na pessoa, no tempo e no modo adequados.

Deve essere arrivato tardi e si vergognava a dirlo.
Deve essere arrivato tardi e vergognava di dirlo.

16. ADVÉRBIOS E LOCUÇÕES ADVERBIAIS

REGRA 16.1

Advérbio *prima*: quando o tempo de referência não ocorre de maneira explícita na frase, substituir *prima* por *con antecedenza*.

Maria dovette uscire prima per non perdere l'aereo.

Maria ha avuto di uscire con antecedenza per non perdere l'aereo.

REGRA 16.2

Locução adverbial *in tempo*: substituir por *a tempo*.

Se arriviamo in tempo andremo a giocare a tennis insieme.

Se arriviamo a tempo andremo a giocare tennis insieme.

REGRA 16.3

Locução adverbial *lo stesso*: Substituir por *nella stessa*.

Benché possa piovere, usciremo lo stesso.

Benché possa piovere usciremo nella stessa.

17. EXPRESSÕES IDIOMÁTICAS

REGRA 17

Expressão *essere in grado*: substituir por *essere capace*.

Penso che non sono in grado di terminare il lavoro.

Penso che non sono capace di terminare il lavoro.

REFERÊNCIAS BIBLIOGRÁFICAS

- ARNOLD, D., BALKAN, L., MEIJER, S., HUMPHREYS, R.L., SADLER, L. (1994), *Machine Translation: an Introductory Guide*, London, NCC Blackwell.
- BANJAR, S.Y. (2001), *Controlled Language and Machine Translation*, Bulletin of the Faculty of arts, volume 17, pp. 34-91.
- BECKER, M., REMBERGER, E. (2010), *Modality and Mood in Romance. Modal interpretation, mood selection, and mood alternation*, Berlin: De Gruyter.
- BHATTACHARYYA, P. (2014), *Machine Learning for Machine Translation*, CSE Dept., IIT Bombay, ISI Kolkata, Jan 6, 2014. Disponível em http://www.isical.ac.in/~acmsc/TMW2014/P_bhattacharyya.pdf.
- BRITO, A.M. (2003), *Subordinação adverbial* in Mateus *et al.* (org.), *Gramática da Língua Portuguesa*, 7ª edição, Lisboa: Editorial Caminho, pp. 695-728.
- CHÉRGOVA, V. (2009), *Particularidades do sistema verbal português no indicativo*, Masarykova, Études Romanes de Brno, 30, 2009, 1.
- CLARK, P. *et al.* (2009), *Naturalness vs. Predictability: a Key Debate in Controlled Languages*, Seattle, Boeing Research and Technology, Proceedings of CNL 2009, pp. 65-81. Disponível em: <http://www.cs.utexas.edu/users/pclark/papers/cnl09.pdf>
- COSTA CAMPOS, M.H. (1998), *Dever e poder: um subsistema modal do Português*, Lisboa: Fundação Calouste Gulbenkian.
- COSTA-JUSSÀ, M. R. *et al.* (2012), *Study and comparison of Rule-based and Statistical Catalan-Spanish machine translation systems*, Computing and Informatics, volume 31, 2012, pp. 245-270.
- COSTA-JUSSÀ, M. R., FONNOLOSA, J.A.R. (2015), *Latest trends in hybrid machine translation and its applications*, Computer Speech & Language, volume 32, Issue 1, jul 2015, pp. 3-10. Disponível em <http://dx.doi.org/10.1016/j.csl.2014.11.001>.
- CREMERS, L. (2011), *Controlled Language and Machine Translation. A practical implementation*, in The big wave, 42.
- CUNHA, C., CINTRA, L. (1998), *Breve gramática do português contemporâneo*, Lisboa: Edições Sá da Costa.

- CZOPEK, N., *Dever e Poder como exemplos de verbos modais nas línguas portuguesa e espanhola*, Kraków, Uniwersytet Jagielloński.
- DA SILVA, J.M. (2012), *Modo, Modalidade, Modalização: Autonomia semântico-discursiva em expressões verbais*, Prolíngua, Volume 7, número 2, jul/dez de 2012, pp. 41-58.
- DARDANO, M., TRIFONE, P. (1995), *Grammatica italiana con nozioni di linguística*, Milano: Zanichelli.
- DORR, B.J., JORDAN, P.W., BENOIT, J.W. (1998), *A Survey of Current Paradigms in Machine Translation*, Institute for Advanced Computer Studies, University of Maryland, College Park.
- DUARTE, I. (2003), *Subordinação completiva – as orações completivas* in Mateus *et al.* (org.), *Gramática da Língua Portuguesa*, 7ª edição, Lisboa: Editorial Caminho, pp. 593-652.
- ELLISTON, J. S. G. (1979), *Computer-aided translation – a business viewpoint*, in *Translating and the Computer: proceeding of a seminar*, London, 14th November 1978, Amsterdam: Ed. BM Snell, pp. 149-158.
- FELLET, A. (2008), *Traduzione automatica: produttività, qualità, customer satisfaction*, Dissertação de Mestrado. Roma, Università LUSPIO.
- HOGEWEG, L., DE HOOP, H., MALCHUKOV, A. (2009), *Cross-linguistic Semantics of Tense, Aspect, and Modality*, Amsterdam: John Benjamins Publishing Company.
- HUIJSEN, W. (1998), *Controlled Language: An introduction*, in Mitamura *et al.* (org.), pp 1-15.
- HUTCHINS, J.W. (2002), *The state of machine translation in Europe and future prospects*. HLT Central, January 2002. Disponível em <http://hutchinsweb.me.uk/HLT-2002.pdf>.
- HUTCHINS, J.W. (2003), *Commercial systems: The state of the art*, *Computers and translation: a translator's guide*, edited by Harold Somers, pp.161-174.
- HUTCHINS, J.W. (2003), *Machine Translation: General Overview*. Disponível em <http://hutchinsweb.me.uk/Mitkov-2003.pdf>.
- HUTCHINS, J.W. (2003), *Machine translation: half a century of research and use*, Prepared for UNED summer school at Ávila, Spain, July 2003. Disponível em <http://www.hutchinsweb.me.uk/Avila-2003.pdf>.

- HUTCHINS, J.W. (2010), *Machine Translation: A Concise History*, Journal of Translation Studies 13, vol. 1-2, Special issue: The teaching of computer-aided translation, Chan Sin Wai (ed.), Chinese University of Hong Kong, pp. 29-70.
- HUTCHINS, J.W. (2014), *Machine translation: History of Research and Applications*, Formerly University of East Anglia, the United Kingdom. Disponível em <http://www.hutchinsweb.me.uk/Routledge-2014.pdf>.
- HUTCHINS, J.W., SOMERS H.L. (1992), *An introduction to Machine Translation*, Cambridge: Academic Press.
- IBRAHIMO, N. (2010), *Para uma Tradução Automática baseada em Conhecimento: especificação da modificação e da predicação adjetival*, Dissertação de Mestrado, Lisboa, Universidade de Lisboa.
- KAMPRATH, C., ADOLPHSON, E., MITAMURA T. & NYBERG, E. (1998), *Controlled Language for Multilingual Document Production: Experience with Caterpillar Technical English*, in Proceedings of the Second International Workshop on Controlled Language Applications, CLAW 98, Pittsburgh, PA, pp. 51–61.
- KUHN, T. (2014), *A Survey and Classification of Controlled Natural Language*, ETH Zurich, University of Zurich, Computational Linguistics, Volume 40, número 1, pp. 121-170.
- MARQUES, R. (1995), *Sobre o valor dos modos conjuntivo e indicativo em português*, Dissertação de Mestrado em Linguística Portuguesa Descritiva, Faculdade de Letras da Universidade de Lisboa.
- MARQUES, R. (2014), *Modalidade e condicionais em português*, ReVEL, edição especial n.8, 2014, pp. 106-130.
- MARRAFA, P. AMARO, R., FREIRE, N., MENDES, S. (2012), *Portuguese Controlled Language: Coping with Ambiguity*, CLG – Group for the Computation of Lexical and Grammatical Knowledge, Centro de Linguística da Universidade de Lisboa.
- MARRAFA, P., AMARO, R., MENDES, S., IBRAHIMO, N. (2011), *Português controlado para tradução automática e para ensino/aprendizagem do português*, CLUL/Instituto Camões. Disponível em <http://www.clul.ul.pt/clg/eng/projectos/portcontrol.html>.
- MATEUS, M.H., BRITO A.M., DUARTE I., HUB FARIA I. (2003), *Gramática da Língua Portuguesa*, 7ª Edição, Lisboa: Editorial Caminho.

- MITAMURA, T. (1999), *Controlled Language for Multilingual Machine Translation*, MT Summit VII, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, USA, sept. 1999, pp.46-52.
- MUÑIZ MUÑIZ, M.N. (1997), “*Esserci”/”Essere” e “haber”/”estar”*: problema contrastivi e proposte didattiche, Italiano & Oltre, nº5, pp. 287-306.
- NIRENBURG, S. (1989), *Knowledge-Based Machine Translation*, Machine Translation 4, pp. 5-24.
- NYBERG, E., MITAMURA, T. (1996), *Controlled English for Knowledge-Based MT: Experience with KANT system*, in Proceedings of the 6th international workshop on theoretical and methodological issues in machine translation (TMI) 95, Leuven, Bélgica, julho 5.7.
- NYBERG, E., MITAMURA, T., HUIJSEN W. (2003), *Controlled language for authoring and translation* in Harold Somers, editor, *Computers and Translation: A Translator’s Guide*. John Benjamins Publishing Company, pp. 245-281.
- OLIVEIRA F. (2001), *Some Issues about the Portuguese Modals dever and poder* in Van Der Auwera-Dendale, pp. 167-184.
- OLIVEIRA, F. (2003), Modalidade e modo, in Mateus *et al.* (org.), *Gramática da Língua Portuguesa*, 7ª edição, Lisboa: Editorial Caminho, pp. 243-272.
- OLIVEIRA, F. (2003), *Tempo e aspecto* in Mateus *et al.* (org.), *Gramática da Língua Portuguesa*, 7ª edição, Lisboa: Editorial Caminho, pp. 127-178.
- PALMER, F.R. (2001), *Mood and Modality*, Cambridge Textbooks in Linguistics.
- PARISI D., ANTINUCCI F., CRISARI M. (1975), *Dovere, potere, volere e il futuro dei verbi*, in Parisi (1975), pp. 238-270.
- PIETRANDREA, P. (2002), *La modalità epistemica. Cornici teoriche e applicazioni all’italiano*, Roma, Università degli studi di Roma Tre.
- PUSTEJOVSKY, J. (1995), *The Generative Lexicon*, Cambridge: MIT Press, Massachusetts.
- REUTHER, U. (2003), *Two in one – can it work? Readability and translatability by means of controlled language*, EAMT-CLAW03, Dublin City University, 15-17 May 2003, pp.124-132.
- ROCCI, A. (2005), *On the nature of the epistemic readings of the Italian modal verbs: the relationship between propositionality and inferential discourse relations*, Lugano, Università della Svizzera Italiana.

- ROTHSTEIN, B., THIEROFF, R. (2010), *Mood in the Languages of Europe*, Wien: John Benjamins Publishing Company.
- SCHWITTER, R. (2002), *Working for two: A bidirectional grammar for a controlled natural language*, in Proceedings of AI 2008, pp. 168-179, Auckland.
- SERIANNI, L. (2010), *Grammatica italiana. Italiano comune e lingua letteraria*, Novara: UTET Università.
- SLOCUM, J. (1984), *Machine Translation: its History, Current Status, and Future Prospects*.
- SQUARTINI, M. (1998), *Verbal periphrases in romance aspect, actionality, and grammaticalization*, Berlin: Mouton de Gruyter.
- TUCCI, I. (2005), *L'espressione della modalità nel parlato: i verbi modali nei corpora italiano e spagnolo C-ORAL-ROM*, in Korzen J., *Lingua cultura e intercultura: l'italiano e le altre lingue*, Atti dell' VIII Convegno Internazionale della Società di Linguistica e Filologia Italiana (SILFI), Samfundslitteratur, Copenhagen, pp. 295-308.
- WEINREICH U. (1964), *Webster's Third: A Critique of his Semantics*, International Journal of American Linguistics, pp. 405-409.
- WU, H., WANG, H. (2009), *Revisiting Pivot Language Approach for Machine Translation*, Toshiba (China), Research and Development Center, Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp 154-162.
- ZAMAGNI, A. (2014), *Italiano Controlado para a Tradução Automática (italiano-português). Linguagem especializada: informática*, Dissertação de Mestrado, Lisboa, Universidade de Lisboa.
- ZUCCARELLO, M.F., BELIZÁRIO, E.S. (2009), *As preposições acidentais (preposizioni improprie) italianas e seus termos correspondentes em português*, in XII Congresso Nacional de Linguística e Filologia, 25-29 Agosto 2008, Rio de Janeiro, Caderno do CNLF, vol. XII, nº16, pp. 70-78.

REFERÊNCIAS SITOGRÁFICAS

- Enciclopedia Treccani <http://www.treccani.it/>
- Priberam <https://www.priberam.pt/>
- Systran <http://www.systran.it/>
- SystraNet <http://www.systranet.com/translate>