ESTUDO PRELIMINAR DA ANÁFORA NUM CORPUS DE INTERPRETAÇÃO SIMULTÂNEA

SÍLVIA ARAÚJO E ANA CORREIA

ABSTRACT

In this paper, we propose an exploratory study about the usefulness of multilingual corpora in areas related to the study of language, translation and, in particular, of simultaneous interpreting. After a brief overview of corpus-based interpreting studies as well as of some existing electronic interpreting corpora, we move on to describe the compilation stages of a bidirectional multimedia corpus (PT–EN/EN–PT). This is followed by an example of how the corpus can be explored, which focuses on the issue of anaphoric relations. The aim of this study is twofold: on the one hand, to convey the relevance of this type of resource as a repository of authentic simultaneous interpreting data; and, on the other hand, to demonstrate that by analysing it from a linguistic perspective it may be possible to identify sensitive areas in simultaneous interpreting (e.g. anaphora), which may prove an important contribution for interpreter training.

[1] LINGUÍSTICA DE CORPUS APLICADA À INTERPRETAÇÃO

Após o comprovado êxito da linguística de corpus aplicada ao estudo da tradução, seguiram-se vários trabalhos de investigação com o intuito de aplicar esta metodologia ao contexto específico da interpretação¹, incidindo maioritariamente nos modos consecutivo e simultâneo, até ao momento. Do conjunto de estudos sobre interpretação baseados em corpora, refira-se, a título de exemplo, o de Timarová (2005) sobre a densidade lexical dos discursos, o de Russo et al. (2006) sobre padrões lexicais e, finalmente, o de Shlesinger (2009) sobre as características ditas universais do discurso interpretado ou *interpretese*. Importa ressaltar que existem, para além dos corpora de interpretação eletrónicos apresentados na secção seguinte, várias compilações manuais (Maia 2000, 2008) de discurso interpretado que também constituem corpora — a diferença reside no facto de não estarem organizados numa plataforma eletrónica. O estudo de Setton (1999), por exemplo, baseia-se num corpus *ad hoc* compilado manualmente. Uma parte significativa destes corpora não informatizados é composta pelas produções orais de alunos a frequentar pós-graduações de interpretação, que são gravadas e posteriormente

[1] O termo interpretação deve ser entendido aqui como a atividade de natureza oral que implica a passagem de uma mensagem de uma língua para outra, quer em modo simultâneo, quer em modo consecutivo (Bendazzoli 2010).

transcritas. Este tipo de coletânea tem vindo a ser utilizado para identificar, com maior ou menor grau de sistematicidade, os erros cometidos pelos alunos, chegando mesmo a culminar na criação de tipologias de erro. A título de exemplo, citamos os estudos de [Gile \(1987, 1992\)](#) e [Falbo \(1998\)](#) sobre os erros cometidos por estudantes de interpretação simultânea durante o seu período de aprendizagem. Estes estudos baseiam-se na premissa de que o trabalho de reflexão sobre a prática da interpretação através de dados reais, i.e., de um corpus (informatizado ou não), promove a consciência metalinguística nos estudantes e pode, por conseguinte, ajudar a desenvolver estratégias de antecipação e resolução de problemas ([Sandrelli 2010](#)). Uma das áreas que pode beneficiar deste tipo de reflexão diz respeito, como veremos na secção 4, às relações anafóricas e às conexões semântico-pragmáticas por elas desencadeadas. Estas podem, de facto, afetar a inteligibilidade do discurso se não forem devidamente processadas, quer ao nível da produção, quer ao nível da receção.

[2] CORPORA DE INTERPRETAÇÃO EM FORMATO ELETRÓNICO

Como vimos, a metodologia da linguística de corpus tem sido usada, ainda que de forma artesanal, para estudar a interpretação. Devido à natureza intrínseca da interpretação, os corpora compilados com material interpretado, manuais ou eletrónicos, são necessariamente paralelos, isto é, contêm pelo menos duas versões linguísticas de um mesmo texto (original e tradução). Daqui se depreende que o corpus paralelo seja de grande utilidade para o estudo da tradução, podendo — e devendo — esta inferência ser alargada ao contexto da interpretação ([Ginezi 2014](#)).

Desde o apelo lançado por [Shlesinger \(1998\)](#), foram desenvolvidos diversos projetos dedicados à compilação de corpora de interpretação eletrónicos, entre os quais se destaca o pioneiro *European Parliament Interpreting Corpus* (EPIC). O EPIC foi indubitavelmente um forte catalisador de investigação nesta área, tendo servido de base a inúmeros projetos de dissertação no contexto académico italiano ([Russo 2010](#)). Os membros da equipa que esteve na origem do EPIC têm-se dedicado ao estudo da direcionalidade e seu impacto no desempenho dos intérpretes, associado às eventuais diferenças decorrentes da interpretação entre línguas românicas, por um lado, e entre uma língua românica e uma germânica, por outro lado ([Monti et al. 2005](#)). Outro estudo que merece destaque resulta do *European Parliament Translation and Interpreting Corpus* (EPTIC), projeto derivado do EPIC, que reequaciona o conceito de simplificação lexical enquanto universal de tradução/interpretação através de medida quantitativas fornecidas pelo corpus ([Bernardini et al. 2013](#)).

Na senda do EPIC ([Bendazzoli & Sandrelli 2005](#)), surgiram outros corpora de interpretação que cobrem uma maior variedade de línguas, modos e contextos de interpretação. É de destacar o trabalho desenvolvido pelo *Hamburger Zentrum*

für Sprachkorpora no âmbito da compilação de corpora orais. No seu repositório, incluem-se também alguns corpora de interpretação, nomeadamente o corpus *Dolmetschen im Krankenhaus* (DiK) (House et al. 2012) e o corpus *Consecutive and Simultaneous Interpreting* (CoSi) (Bührig et al. 2012). O primeiro visa a interpretação comunitária num contexto hospitalar e o segundo está direcionado para a interpretação em situação de conferência, quer em modo consecutivo quer em modo simultâneo. Os três corpora mencionados resultaram de projetos financiados de larga escala. Existem outros, resultantes de projetos individuais de investigação, que constituem também importantes ferramentas para o estudo da interpretação. Um exemplo é o corpus *DIRSI-C*, composto por discursos de natureza médica em inglês e italiano, que permitiu chegar a importantes conclusões sobre o impacto da direcionalidade (i.e., interpretar de ou para a língua materna) no desempenho dos intérpretes (Bendazzoli 2010).

[3] CORPUS DE INTERPRETAÇÃO/PER-FIDE

O corpus de interpretação que constitui a base empírica do estudo sobre a anáfora que apresentamos na secção seguinte decorre do corpus *Per-Fide* (Araújo et al. 2010; Almeida et al. 2014). Para dar continuidade a este corpus composto exclusivamente por textos (escritos) paralelos, pretende-se agora acrescentar uma dimensão oral através da compilação de um corpus de interpretação. Este corpus será composto pelas transcrições das intervenções, em sessão plenária, dos eurodeputados portugueses e ingleses. Convém, aqui, realçar que as transcrições por nós realizadas são diferentes das transcrições que integram o corpus *Europarl* (Koehn 2005). Com efeito, para o corpus de interpretação, partimos do *Compte-Rendu in Extenso* (CRE), i.e., o relato integral das sessões plenárias do Parlamento Europeu, que deu origem ao *Europarl*, mas procurámos aproximá-lo daquilo que é efetivamente proferido pelos deputados e intérpretes. Apesar de se limitar às línguas portuguesa e inglesa, este corpus de interpretação tem a vantagem de especificar a língua-fonte de cada intervenção.

Além de ser um produto derivado do projeto *Per-Fide*, este corpus que agora apresentamos é diretamente influenciado pelo *EPIC*, projeto que lhe serviu de inspiração. De seguida, passamos então a descrever o Corpus de Interpretação/*Per-Fide*, salientando algumas das suas características e elencando as várias etapas de compilação, sem nos adentrarmos nos detalhes mais técnicos. Porém, antes ainda dessa descrição, apresentamos um resumo visual do processo, que poderá contribuir para uma melhor compreensão dos procedimentos envolvidos na compilação do corpus na figura 1.

Conforme assinalado na figura 1, a criação do corpus pode ser decomposta em três grandes etapas: pré-processamento, processamento e pós-processamento. Antes de o material a incluir no corpus poder ser processado, é necessário levar a cabo a transcrição do material, ainda em formato audiovisual (pré-processamento).

FIGURA 1: Etapas de compilação do corpus

De seguida, é necessário segmentar, alinhar e anotar esses dados transcritos (processamento), para que possam ser depois pesquisados na rede através de uma interface construída para o efeito (pós-processamento). Abaixo, apresentamos estas três fases de forma um pouco mais detalhada.

Pré-processamento

Conforme mencionado acima, o Corpus de Interpretação/*Per-Fide* é um corpus de interpretação composto pelas transcrições dos discursos proferidos pelos euro-deputados portugueses e ingleses e das respectivas interpretações. Os discursos reunidos correspondem a um período de seis meses de intervenções dos deputados portugueses, desde janeiro a junho de 2011. A fim de equilibrar quantitativamente os dados, as intervenções dos homólogos ingleses limitam-se a apenas três desses seis meses. Este é um corpus bidirecional na medida em que inclui discursos originais em ambas as línguas. Podemos, aliás, dividi-lo em dois subcorpora para melhor ilustrar o seu carácter bidirecional: subcorpus 1) português original — inglês interpretado; subcorpus 2) inglês original — português interpretado. Cria-se, assim, uma estrutura cruzada que, como se pode ver abaixo na figura 2, demonstra a natureza simultaneamente paralela e comparável (a dois níveis) do corpus.

Os discursos (i.e., originais e interpretações) que integram este corpus apresentam especificidades decorrentes do contexto em que ocorrem. Com efeito, o sistema de atribuição da palavra no Parlamento Europeu é bastante rígido, o que faz com que cada deputado tenha, em média, dois minutos para as suas intervenções. Para garantir a máxima rentabilização de um tempo de antena mínimo, os

FIGURA 2: Esquema de combinatórias linguísticas.

deputados leem os seus discursos, que são previamente redigidos. Naturalmente, a leitura dos discursos traduz-se num débito e densidade elevados, que tipicamente não se encontrariam numa produção espontânea. No que toca aos intérpretes, não sabemos se estes têm acesso prévio aos discursos que vão interpretar, assim como não dispomos de informação sobre a sua formação e experiência profissional.

Os discursos que vão integrar o corpus foram descarregados em bruto a partir do sítio web do serviço audiovisual do Parlamento Europeu, que disponibiliza as gravações das sessões plenárias desde abril de 2006 em todas as línguas oficiais da União Europeia. Estes discursos foram posteriormente armazenados no servidor do projeto *Per-Fide*. Cada discurso corresponde a um ficheiro *wmv*, que por sua vez integra 22 pistas de áudio (uma para cada língua oficial, à exceção do croata e do irlandês). Assim, o passo seguinte consistiu em isolar para cada vídeo o material áudio relevante, i.e., as pistas portuguesa e inglesa. No que respeita aos direitos de autor, importa referir que o serviço audiovisual permite que o material disponibilizado no seu sítio web seja utilizado gratuitamente para fins educativos, conforme se pode ler no separador *Copyright* desse mesmo sítio.

Depois de recolher e processar os discursos, deu-se início à etapa de transcrição. Contámos, para isso, com o apoio dos alunos do 3º ano da licenciatura em Línguas Aplicadas (2013/14) da Universidade do Minho que transcreveram, no âmbito da unidade curricular de Princípios de Interpretação, a quase totalidade dos discursos a integrar no corpus. Nesta etapa, o investigador é confrontado com a necessidade de tomar decisões metodológicas que orientem o trabalho de transcrição. Estas decisões devem estar firmadas na identificação clara da(s) fi-

nalidade(s) do corpus. No nosso caso, por exemplo, o esforço adicional de anotar pausas e hesitações ao transcrever os discursos não acrescentaria valor ao estudo das relações anafóricas, que é o fenómeno que pretendemos estudar através deste corpus. Contudo, devemos ressaltar que, embora concebido para o estudo de um fenómeno particular, este corpus não se limita ao estudo da anáfora. Será um recurso aberto à comunidade académica, podendo servir de base a inúmeras pesquisas sobre aspetos linguísticos e cognitivos da linguagem em uso. O trabalho de transcrição pautou-se por uma preocupação constante em refletir o mais fielmente possível aquilo que é efetivamente pronunciado pelos oradores e intérpretes. Assim, a par da transcrição ortográfica, optou-se por anotar casos de autocorreção e reformulação, utilizando uma barra (/). As transcrições basearam-se nas convenções HIAT (*Halbinterpretative Arbeitstranskriptionen*), integradas no EXMARaLDA Partitur-Editor (Schmidt 2004). HIAT é um sistema de transcrição para a representação escrita da língua falada, projetado nos anos 70 no âmbito da análise do discurso e que foi retomado pelos criadores do Partitur-Editor EXMARaLDA. As transcrições propriamente ditas foram elaboradas no referido *software*, Partitur-Editor, que permitiu ainda a integração de metadados relativos aos oradores e aos discursos através dos separadores *Speakertable* e *Metainformation*, respetivamente. No primeiro separador, inclui-se o nome do deputado e o grupo político ou o sexo do intérprete (M/F), conforme se trate de um discurso original ou interpretação. No segundo separador, inclui-se a data do discurso e a fonte (i.e., original vs. interpretação). A escolha deste *software* foi determinada pelo seu carácter livre e, sobretudo, pelo facto de codificar os ficheiros transcritos em XML. Cabe assinalar que o EPIC, tratando-se de um projeto com mais recursos humanos e financeiros, apresenta transcrições mais detalhadas do ponto de vista da anotação paralinguística (e.g. pausas e disfluências). Note-se que a tentativa de produzir transcrições que sejam *user/annotator-friendly* é comum a ambos os corpora. O nível de detalhe é também superior no que concerne à anotação paralinguística, já que cada transcrição é acoplada a um cabeçalho de metadados que inclui abundante informação sobre o contexto, orador e o discurso (e.g. tema, tema específico, duração, número de palavras, palavras por minuto, etc.). A inclusão de metadados, também prevista no nosso corpus, garante que outros investigadores possam utilizar estes recursos de uma forma adequada, já que a origem e a natureza dos dados estão devidamente identificadas.

Processamento

Como se disse, neste momento, encontramos-nos em fase de revisão das transcrições já elaboradas pelos alunos da licenciatura em Línguas Aplicadas. Após esta etapa, passar-se-á a um conjunto de procedimentos técnicos que visam a estruturação e disponibilização do corpus na rede, em livre acesso. Feitas as transcrições, será necessário alinhar os originais com as respetivas transcrições. Este

alinhamento implica uma segmentação ao nível frásico, passível de automatização graças aos códigos XML gerados pelo *Partitur-Editor*. Esta segmentação servirá de base ao alinhamento dos bi-textos (i.e., original + interpretação) mas também à própria segmentação dos ficheiros vídeo/áudio correspondentes que serão depois sincronizados com o texto. Daqui decorre o carácter multimédia deste corpus, à semelhança do corpus multimédia *Veiga* de legendagem (Dios & Guinovart 2012), desenvolvido no *Centro de Linguística da Universidade de Vigo*. Este corpus irá, assim, permitir ao utilizador efetuar pesquisas em bitextos alinhados, com acesso ao material audiovisual correspondente. Com efeito, estes dois aspetos representam um salto evolutivo face ao *EPIC*, embora, no âmbito deste último, esteja já prevista a inclusão de bitextos alinhados e de material audiovisual como sugestão de trabalho futuro. É importante referir que os bitextos serão alinhados e etiquetados morfossintaticamente, aumentando as potencialidades do corpus para fins de investigação. Estas duas tarefas serão executadas de forma semiautomática, combinando *software* de alinhamento (Simões & Almeida 2007) e etiquetação (Schmid 1994; Brants 2000) cujos resultados serão alvo de uma revisão manual.

Pós-processamento

A última etapa prende-se com a disponibilização do corpus e prevê a construção de uma interface de pesquisa própria, que a seu tempo poderá ser consultada através do sítio web do corpus *Per-Fide*.

Neste momento, o corpus ainda não se encontra disponível para consulta. Contudo, foi possível analisar um subconjunto de discursos, já transcritos, que farão parte do corpus. Esta análise revelou que a anáfora é um fenómeno linguístico importante na interpretação, uma vez que pode afetar os discursos produzidos pelos intérpretes em termos de coesão e coerência.

[4] A ANÁFORA NA INTERPRETAÇÃO SIMULTÂNEA

Toda a comunicação significativa usa a linguagem e, se quisermos saber como esta funciona, temos que analisar atos concretos de comunicação, i.e., textos (orais e/ou escritos). Compreender um texto implica ser capaz de construir uma representação mental coerente desse mesmo texto (Morais 2011, pg. 17). Como refere (Lopes 2008, pg. 62), há duas dimensões que se afiguram cruciais na construção de uma representação mental de um texto. Em primeiro lugar, a *coerência referencial*, basicamente suportada pelas anáforas discursivas e, em segundo lugar, a *coerência relacional* que envolve relações semânticas que conectam segmentos discursivos. São muitos os trabalhos dedicados a esta dimensão da coerência discursiva, que permite estabelecer uma fronteira entre uma sequência aleatória e desconexa de frases e um produto com sentido em termos semântico-pragmáticos (Jaubert 2005; Kostopoulou 2007; Sanders et al. 2007; Lopes & Rodrigues 2013, entre muitos outros). Ao produzir um discurso, um determinado locutor utiliza mecanismos de

coesão e coerência que lhe permitem assinalar a interdependência semântica entre os diferentes elementos que o compõem. Tanto quanto é do nosso conhecimento, não foi ainda realizado um estudo especificamente centrado na compreensão da anáfora e ancorado em dados extraídos de um corpus de interpretação simultânea. Nas linhas que se seguem, pretendemos demonstrar que os processos anafóricos são pistas fortes para a ativação e orientação das operações cognitivas conducentes à compreensão do discurso original e da sua interpretação.

[4.1] *A relação anafórica: uma relação binária assimétrica*

Uma relação anafórica caracteriza-se pela dependência interpretativa entre dois termos ou segmentos textuais, o *termo antecedente* (ou termo anaforizado) e o *termo anafórico* (ou termo anaforizante). A anáfora é pois uma relação binária assimétrica entre um antecedente previamente identificado referencialmente e um termo anafórico que só ganha identidade referencial pela sua integração naquele contexto (Campos & Xavier 1991, entre outros). A título meramente exemplificativo, podemos dizer que o excerto do discurso original que se segue entrelaça duas cadeias de referência (uma cadeia com três anéis — *excelente trabalho...que (efectuou)...que (se traduz)* e uma outra cadeia com quatro anéis — *relatório...que (permite)... elipse de que (constitui)...este (é um relatório)*):

- (1a) Muito obrigado, Senhor Presidente, caros Colegas. Quero felicitar os relatores pelo **excelente trabalho** que efectuou e que se traduz num **relatório** que permite dar um bom início à construção do próximo quadro financeiro plurianual e [que] constitui um desafio para a Comissão e para o Conselho. Este é um relatório ambicioso e em simultâneo um relatório realista.

Cada uma destas cadeias referenciais faz uso de procedimentos léxico-gramaticais (Oliveira 1987; Kleiber 1994; Corblin 1995; Perdicoyanni-Paléologou 2001, entre outros) que instauram a dependência de um conjunto de termos anafóricos relativamente a uma expressão nominal plena que inicia a cadeia referencial. Todos estes termos anafóricos pronominais retiram a sua referência exclusivamente da relação que têm com o seu antecedente. É interessante notar que, ao restituir este discurso original, o intérprete inglês elimina a marca de plural (*the rapporteur*) que o deputado aplica por engano no actante agentivo (*os relatores*) do verbo *efectuou* acima transcrito:

- (1b) Yes, colleagues, I would like to congratulate **the rapporteur** for his excellent work because he permits us to really do things in the right way. A good start, financially speaking, within the framework of the MFF and of course what we have here is a very ambitious report, but a very realistic one as well.

Neste discurso interpretado, mantém-se a significação global do discurso original, apesar de a cadeia de referência mais saliente ((*the rapporteur*)...*his (excellent work)*...*he (permits)*) incidir sobre a expressão nominal *the rapporteur* e já não sobre os termos antecedentes estipulados em (1a). Ou seja, o discurso interpretado deixa de dar primazia ao produto (i.e., ao relatório) para abrir com o autor desse produto (i.e., o relator) a cadeia de referência que percorre o excerto (1b). Trata-se, na realidade, de uma restituição da informação por modulação metonímica (Chuquet & Paillard 1987, pg. 31) que consiste em privilegiar a relação de causa (*rapporteur*) pelo efeito (*report*) e não o inverso.

A referência anafórica é, sem dúvida, uma condição básica para a construção de qualquer ato comunicativo. É ela que contribui para a organização textual, na medida em que assegura a progressão temática. Importa assinalar, contudo, que o uso excessivo de elementos anafóricos pode constituir um obstáculo à clareza. De facto, os deputados portugueses tendem a alongar as frases através de mecanismos recorrentes de subordinação (com o uso do pronome relativo *que*), conforme ilustrado no exemplo (2a):

- (2a) **O longo processo de trabalho que este importante relatório exigiu**, incluindo os muitos compromissos alcançados, tornou-o num documento bastante amplo e equilibrado dos **diversos interesses que a PAC tem de dar resposta**. Este relatório constitui uma boa orientação para as propostas legislativas, pelo que felicito o seu relator.

Esta concatenação de *que* exige um esforço cognitivo adicional e maior tempo de processamento, dado que frequentemente está associada a verbos com diferentes tipos de transitividade. Por exemplo, a colocação *dar resposta* mencionado em (2a) seleciona um complemento introduzido por intermédio da preposição *a*. Mas este traço sintático nem sempre é respeitado, sobretudo com alguns complementos de natureza oracional, nomeadamente nas orações relativas, como podemos ver no exemplo anterior ('dos diversos interesses que a PAC tem de dar resposta' em vez de 'dos diversos interesses a que a PAC tem de dar resposta'). Nestes casos, a omissão da preposição tem vindo a generalizar-se. Para adaptar este discurso aos moldes sintáticos da língua inglesa, o intérprete opta por dividir este parágrafo em orações coordenadas assindéticas ('we've been through [...]'; 'there's been many [...]'; 'now we have [...]') ou sindéticas ('and I would like [...]') de forma a evitar a cascata de pronomes relativos do discurso original:

- (2b) **We've been through a long procedure, there's been many amendments that have been tabled and compromises reached. Now **we** have a balanced report taking into account different interests. This report is a good guideline for proposals, legislative proposals, and I would like to congratulate the rapporteur.**

Ao optar maioritariamente por um esquema paratático (Duarte 2003) que consiste em verter o conteúdo das orações relativas dentro de orações coordenadas (as)sindéticas, o intérprete é necessariamente levado a segmentar o seu discurso em blocos de informação mais circunscritos que preservam a ordem canónica da frase (sujeito — verbo — objeto(s)). Esta forma aparentemente mais simples (porque mais atomizada) de processar e restituir o discurso original faz com que o intérprete tenha de atribuir ao verbo de cada uma das orações coordenadas um sujeito, pois o inglês distingue-se do português por ser necessariamente uma língua de sujeito foneticamente realizado. Em (2b), os dois pronomes pessoais *we* são suficientemente impessoais para que o intérprete possa atribuir, sem se comprometer, um sujeito sintático a cada uma das orações independentes que constituem o seu discurso. Uma alternativa à introdução destes pronomes poderia ser o recurso a estruturas passivas curtas que dispensam precisamente qualquer explicitação do agente do processo. É isso que acontece num dos segmentos discursivos presentes em (2b): ‘there’s been many amendments that *have been tabled and compromises reached*’. A mobilização de esquemas diatéticos impessoais (quer ativos quer passivos) pode implicar uma menor sobrecarga da memória, pois, como se pode ver, por exemplo, no discurso interpretado que se segue, o uso do pronome *se* evita que se tenha de explicitar os agentes mencionados no original (a saber: ‘the UK media’/‘reporters’):

- (3a) UK law is being made in Strasbourg, a monster red-tape factory is closing industry, now we have tax, but because it isn’t football, the people aren’t informed. **The UK media** avoid EU reality, brains are fed with TV soaps, **reporters** say the EU does not affect local issues. The EU is not news.
- (3b) Mas porque não é futebol, as pessoas não estão informadas. Foge-se à realidade. Diz-se que a União Europeia não afeta as questões locais. A União Europeia não faz parte das notícias.

Como vimos, os termos de uma relação anafórica surgem, na linearidade do texto, como uma cadeia anafórica: o termo antecedente é referencialmente independente do termo anafórico, ao passo que este é referencialmente dependente do termo antecedente. Importa referir que a escolha errada de um destes termos no discurso interpretado pode originar alterações de sentido mais ou menos profundas. Apresentamos abaixo dois pares de exemplos que ilustram, respetivamente, uma identificação errónea do termo antecedente e do termo anafórico.

Identificação errónea do termo antecedente

No discurso original que se segue, a sigla *AU* (precipitadamente proferida antes do seu equivalente por extenso: *African Union*) foi apreendida pelo intérprete como sendo a sigla referente à União Europeia. Esta confusão entre *AU* e *EU* terá levado

o intérprete a introduzir no seu discurso uma cadeia de referência relativa à União Europeia que não figura de todo no discurso original:

- (4a) **The AU, the African Union**, could do far more. We have heard many platitudes from the AU but we've seen little concrete action so far.
- (4b) **A União Europeia, a União Africana** com certeza que poderiam fazer mais. Ouvimos já belas declarações da UE mas até agora poucas acções.

A análise deste exemplo exhibe uma conexão interfrásica bem-sucedida do ponto de vista da coesão e da coerência. Contudo, em (4b), o antecedente único do discurso original ('a União Africana') passa a assumir uma entidade dupla ('a União Europeia' e a 'União Africana') no discurso interpretado, que despoleta a retoma parcial da entidade errada, devido provavelmente à semelhança fonética entre 'AU' e 'EU'.

Identificação errónea do termo anafórico

Em português, o termo anafórico pronominal pode ser um pronome pessoal de 3ª pessoa, sujeito ou complemento, com ou sem realização lexical. Quando, na relação anafórica, o termo anafórico não tem realização lexical, i.e., corresponde a uma categoria vazia, falamos de elipse e de termo elíptico (Marques 2009, pg. 38). No discurso original que se segue, estamos perante um sujeito nulo subentendido, que, apesar de não estar foneticamente realizado, pode ser identificado pelas marcas de concordância verbal ('eles] trabalharam bem'):

- (5a) Desde a sua adesão à União, em 2007, que **quer a Bulgária, quer a Roménia** tinham a expectativa legítima dos seus cidadãos se tornarem cidadãos comunitários de pleno direito e poderem usufruir dos mesmos direitos de todos os outros cidadãos comunitários, onde se inclui a liberdade de circulação no interior do Espaço Schengen. É, pois, a cidadania europeia que reforçamos ao alargar o Espaço Schengen. Sexta e última nota, Senhor Presidente: [eles] **trabalharam bem**. É evidente que ambos os países estão de parabéns pelo esforço que realizaram para cumprir todos os requisitos de Schengen.

A interpretação referencial desse pronome [eles] que não está materialmente expresso depende exclusivamente da sua relação anafórica com o SN ('quer a Bulgária, quer a Roménia') que figura no contexto linguístico à esquerda. No discurso interpretado, assistimos a uma quebra desta progressão temática (antecedente: 'Bulgária e Roménia'...anafórico: [pro=eles]), pois o intérprete opta por um pronome pessoal de 2ª pessoa (*you've done a good job*) quando se estaria à espera do pronome de 3ª pessoa (*they've done...*):

- (5b) Since their accession (to the) European Union, **Bulgaria and Romania** have legitimately wanted their citizens to become European citizens, with full rights, and that of course includes the right to move freely within the Schengen area. So European citizenship is what we are about strengthening here by enlarging the/ this Schengen area. Sixth point: **you**'ve done a good job. It's clearly that/ it's clear that these countries should be congratulated for all the efforts they've made to comply with the Schengen requirements, [...]

Neste caso, não parece possível a identificação anafórica de *you* em relação ao SN *Bulgaria and Romania*, a não ser que a situação extralinguística permita identificar referencialmente esse SN. Não parece ser o caso, pois o deputado não interpela diretamente nenhum representante passível de incarnar esse *you*.

Para além dos casos anteriormente mencionados de escolha inadequada de um dos termos da relação anafórica, também detetámos casos de não explicitação do termo anafórico e, como veremos mais à frente, de dupla identificação do termo anafórico.

Não identificação do termo anafórico

O discurso original que se segue contém uma relação anafórica, na qual o SN *as devidas ilações* é o antecedente e o pronome indefinido *uma* é o termo anafórico. Este pronome funciona simultaneamente como termo catafórico ([Perdicoyanni-Paléologou 2001](#)) que anuncia a oração que figura posteriormente na linearidade textual:

- (6a) Espero que todas as instituições tenham retirado **as devidas ilações** da crise financeira, económica e social que teima em não nos largar. **Uma** para mim é clara: só uma Europa unida, solidária, com uma governação económica reforçada, concertada e convergente, estará à altura de enfrentar com sucesso os desafios políticos que temos pela frente.

No discurso interpretado, perde-se esta dupla cadeia referencial (anafórica e catafórica):

- (6b) So I think that all of the institutions have learned their lesson from the major financial crisis that we've experienced. We are a united Europe, a Europe of solidarity and we are trying to converge, to converge our policies.

O marcador de negação restritiva que figura, em (6a), no último elemento da cadeia referencial ('*só uma Europa unida...estará à altura...*') parece-nos importante para reforçar a ideia de que a ilação apresentada resume, de facto, a principal lição que se deve reter e aplicar para se conquistar uma Europa melhor. Ao eliminar

do seu discurso este marcador e ao optar por uma forma verbal no presente simples ('are'), o intérprete não transmite exatamente o mesmo sentido que subjaz ao original e leva-nos a crer que a Europa já 'aprendeu a lição'. Aqui, fica claro que a rutura da cadeia anafórica original tem repercussões, a nível semântico, no discurso-alvo.

A alteração de sentido decorrente dessa rutura poderia ter sido atenuada se a forma verbal 'are' ('We are a united Europe' → oração 1) aparecesse precedida, por exemplo, de um verbo como *need* ('We need to be a united Europe') que marca uma 'necessidade' implicitamente expressa no enunciado original: 'para que a Europa possa estar à altura dos seus ideais, *precisamos* de a tornar mais coesa'. Imposta pela situação de crise que se faz sentir, esta necessidade de implementação de uma (maior) concertação económica devia suscitar uma nova forma de atuação por parte dos agentes europeus. O verbo *try* usado na forma progressiva ('we are trying to converge, to converge our policies' → oração 2) denota, sem dúvida, o esforço que tem sido feito nesse sentido. No entanto, a junção das duas orações resulta, devido aos múltiplos valores que a conjunção *and* pode adquirir em contexto, numa ambiguidade semântica entre a leitura de finalidade ('somos uma europa unida e para tal, estamos a tentar atuar de forma mais concertada') e de causalidade ('como somos uma Europa unida, estamos a tentar atuar de forma mais concertada'). Por via desta operação, perde-se o valor de condição expresso no original ('a Europa só estará à altura se estiver unida'), cuja concretização se situa no futuro, ao contrário das leituras semânticas acima descritas, ambas ancoradas no presente ('are'/'are trying').

Dupla identificação do termo anafórico

Ao analisar o discurso original que se segue e a respetiva interpretação, deparamo-nos com um caso bastante curioso de quebra da progressão temática. Parece-nos, de facto, que a não especificação do montante atribuído pela UE à Grécia em (7b) pode levar-nos a interpretar o SN *novo financiamento equivalente* não à luz desse montante (como seria expectável) mas sim à luz do sintagma preposicional '(financiar) de forma inútil':

- (7a) Everyone can now see that a default in Greece is coming, except the euro zone finance ministers who, 13 months after uselessly committing **110 billion euros**, now seem set to commit a further sum almost as large.
- (7b) Hoje todos nós podemos ver quais são as consequências para a economia grega, excepto os ministros das finanças da União Europeia que após terem **financiado de forma inútil** a economia grega estão dispostas a fazer um novo financiamento equivalente.

Ou seja, ao passo que o adjetivo *large* (em ‘*a further sum almost as large*’) apresenta, em (7a), um valor indubitavelmente quantitativo (que remete para a ideia de que o novo financiamento poderá aproximar-se dos 110 mil milhões de euros concedidos aquando do financiamento anterior), no caso de (7b), o adjetivo *equivalente* (em ‘*um novo financiamento equivalente*’) parece adquirir um valor predominantemente qualitativo, dado que o único termo antecedente disponível no contexto anterior remete precisamente para o sintagma preposicional acima mencionado. Ao enveredarmos por uma leitura deste tipo, apenas sobressai a ideia de que este novo resgate não ajudará a resolver a situação financeira de um país como a Grécia. Para reforçar a dimensão quantitativa do adjetivo *equivalente*, uma solução possível passaria por substituir este adjetivo por um modificador do tipo: ‘um novo financiamento *quase tão avultado quanto o anterior*’.

[5] CONSIDERAÇÕES FINAIS

Apesar de nos encontrarmos ainda na fase de pré-processamento do corpus, foi já possível isolar alguns exemplos para estudar o fenómeno das relações anafóricas no contexto da interpretação simultânea. Com efeito, a anáfora é um mecanismo que condiciona a coerência textual, o que se reveste de especial importância numa atividade como a interpretação, cujo objetivo último consiste em promover a inteligibilidade comunicativa de uma mensagem-fonte. Através deste estudo exploratório, foi possível identificar as quebras de ligação que podem ocorrer numa cadeia anafórica aquando do ato interpretativo. Como pudemos constatar, estas quebras podem igualmente afetar, ainda que em menor grau, a tessitura do discurso original. Os problemas decorrentes da quebra da cadeia anafórica, que envolve, como vimos, dois termos, podem ser atribuídos à identificação errónea do termo antecedente ou anafórico; à não identificação do termo anafórico; e à dupla identificação do termo anafórico, que gera ambiguidade na interpretação do enunciado. Como se pode concluir, a análise de um conjunto reduzido de exemplos permitiu extrair indícios de padrões de comportamento anafórico, que serão alvo de um estudo qualitativo mais aprofundado e complementado com dados numéricos logo que o corpus esteja pesquisável. A par da anáfora, prevê-se que a exploração do corpus possa também apontar para a existência de áreas críticas na interpretação do ponto de vista linguístico. Graças à integração de material textual e audiovisual neste corpus, os investigadores terão à sua disposição um recurso multimédia que permitirá a realização de estudos linguísticos, não só de natureza segmental mas também suprasegmental (e.g. prosódia). Acreditamos que este tipo de reflexão, baseada em corpora, poderá ser uma fonte de contributos válidos para enriquecer os conteúdos formativos na área da interpretação. Estes poderão também ter aplicações relevantes ao nível da tradução e do ensino de línguas.

AGRADECIMENTOS

Este trabalho foi realizado com o apoio da Bolsa de Investigação com a referência SFRH / BD / 88142 / 2012, financiada pela Fundação para a Ciência e Tecnologia no âmbito do Programa Operacional Potencial Humano inscrito no Quadro de Referência Estratégico Nacional (Formação Avançada), participado pelo Fundo Social Europeu e por fundos nacionais do Ministério da Educação e Ciência.

REFERÊNCIAS

- Almeida, José João, Sílvia Araújo, Nuno Carvalho, Idalete Dias, Ana Oliveira, André Santos & Alberto Simões. 2014. The Per-Fide corpus: a new resource for corpus-based terminology, contrastive linguistics and translation studies. Em Tony Berber Sardinha & Telma São Bento Ferreira (eds.), *Working with Portuguese Corpora*, 177–200. Bloomsbury Academic.
- Araújo, Sílvia, José João Almeida, Alberto Simões & Idalete Dias. 2010. Apresentação do projecto Per-Fide: Paralelizando o Português com seis outras línguas. *Linguamática* 2(2). 71–74.
- Bendazzoli, Claudio. 2010. *Il corpus DIRSI: creazione e sviluppo di un corpus elettronico per lo studio della direzionalità in interpretazione simultanea*: Alma Mater Studiorum Università di Bologna. Tese de Doutoramento.
- Bendazzoli, Claudio & Annalisa Sandrelli. 2005. An approach to corpus-based interpreting studies: Developing EPIC (European Parliament Interpreting Corpus). Em Heidrun Gerzymisch-Arbogast & Sandra Nauert (eds.), *MuTra – Challenges of Multidimensional Translation: Conference proceedings*, 1–12.
- Bernardini, Silvia, Adriano Ferraresi & Maja Miličević. 2013. From EPIC to EPTIC: building and using an intermodal corpus of translated and interpreted texts. Apresentação na *46th Annual Meeting of the Societas Linguistica Europea (SLE 2013)*.
- Brants, Thorsten. 2000. TnT – a statistical part-of-speech tagger. Em *6th Applied NLP Conference, ANLP-2000*, 224–231.
- Bühlig, Kristin, Ortrun Kliche, Birte Pawlak & Bernd Meyer. 2012. The corpus “Interpreting in Hospitals”: Possible applications for research and communication training. Em Thomas Schmidt & Kai Wörner (eds.), *Multilingual Corpora and Multilingual Corpus Analysis. Hamburg Studies in Multilingualism* (14), 305–315. John Benjamins.
- Campos, Maria Henriqueta Costa & Maria Francisca Xavier. 1991. *Sintaxe e Semântica do Português*. Universidade Aberta.

- Chuquet, Hélène & Michel Paillard. 1987. *Approche linguistique des problèmes de traduction anglais - français*. Ophrys.
- Corblin, Francis. 1995. *Les Formes de Reprise dans le Discours. Anaphores et Chaînes de Référence*. Presses Universitaires de Rennes.
- Dios, Patricia Sotelo & Xavier Gómez Guinovart. 2012. A multimedia parallel corpus of english- galician film subtitling. Em Alberto Simões, Ricardo Queirós & Daniela da Cruz (eds.), *st symposium on languages, applications and technologies*, 255–266.
- Duarte, Inês. 2003. Aspectos linguísticos da organização textual. Em Maria Helena Mira Mateus, Ana Maria Brito, Inês Duarte & Isabel Hub Faria (eds.), *Gramática da língua portuguesa*, 87–123. Editorial Caminho.
- Falbo, Caterina. 1998. Analyse des erreurs en interprétation simultanée. *The Interpreters' Newsletter* 8. 107–120.
- Gile, Daniel. 1987. Les exercices d'interprétation et la dégradation du français: une étude de cas. *META* 32(4). 420–428.
- Gile, Daniel. 1992. Les fautes de traduction: une analyse pédagogique. *META* 37(2). 251–262.
- Ginezi, Luciana Latarini. 2014. Desafios para a construção de um corpus de aprendizes de interpretação simultânea. *TradTerm* 23. 165–191.
- House, Julianne, Bernd Meyer & Thomas Schmidt. 2012. CoSi - A Corpus of Consecutive and Simultaneous Interpreting. Em Thomas Schmidt & Kai Wörner (eds.), *Multilingual Corpora and Multilingual Corpus Analysis Hamburg Studies in Multilingualism* (14), 295–304. John Benjamins.
- Jaubert, Anna. 2005. *Cohésion et cohérence. Études de linguistique textuelle*. ENS éditions.
- Kleiber, Georges. 1994. *Anaphores et pronoms*. Duculot.
- Koehn, Philipp. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. Em *Conference Proceedings: the tenth Machine Translation Summit*, 79–86.
- Kostopoulou, Georgia. 2007. The role of coherence in text approaching and comprehension: Applications in translation didactics. *Meta* 52(1). 146–155.
- Lopes, Ana Cristina Macário. 2008. Texto, gramática e processamento. Em Fátima Oliveira & Isabel Margarida Duarte (eds.), *O Fascínio da Linguagem - Actas do Colóquio de Homenagem a Fernanda Irene Fonseca*, 57–68.

- Lopes, Ana Cristina Macário & Conceição Carapinha Rodrigues. 2013. *Texto, coesão e coerência*. Almedina.
- Maia, Belinda. 2000. Making corpora: a learning process. Em Silvia Bernardini & Federico Zanettin (eds.), *I corpora nella didattica della traduzione: Corpus Use and Learning to Translate*, 47–60. Cooperativa Libreria Universitaria Editrice Bologna.
- Maia, Belinda. 2008. Corpógrafo. Presentation at TaLC at TaLC: Teaching and Linguistics (Portuguese language) Corpora. <http://www.linguateca.pt/documentos/MaiaWorkshopTaLC2008.pdf>.
- Marques, Isilda Gaspar. 2009. *Anáfora associativa - propostas de abordagem em contexto escolar*: Faculdade de Letras da Universidade de Coimbra. Tese de Mestrado.
- Monti, Cristina, Claudio Bendazzoli, Annalisa Sandrelli & Mariachiara Russo. 2005. Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus). *META* 50(4). s/pp.
- Morais, Maria da Felicidade Araújo. 2011. *Marcadores da estruturação textual: elementos para a descrição do papel dos Marcadores Discursivos no processamento cognitivo do texto*. Centro de Estudos em Letras. Universidade de Trás-os-Montes e Alto Douro. Coleção Linguística 6.
- Oliveira, Fátima. 1987. Cadeias anafóricas: que referência? *Revista da Faculdade de Letras: Línguas e Literaturas, II série* 4. 125–136.
- Perdicoyanni-Paléologou, Hélène. 2001. Le concept d'anaphore, de cataphore et de déixis en linguistique française. *Revue québécoise de linguistique* 29(2). 55–77.
- Russo, Mariachiara. 2010. Reflecting on interpreting practice: graduation theses based on the European Parliament Interpreting Corpus (EPIC). Em Lew Zybatow (ed.), *Translationswissenschaft-Stand und Perspektiven, Innsbrucker Ringvorlesungen zur Translationswissenschaft VI, (vol 12)*, 35–50. Peter Lang.
- Russo, Mariachiara, Claudio Bendazzoli & Annalisa Sandrelli. 2006. Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: extended analysis of EPIC (European Parliament Interpreting Corpus). *Forum* 4(1). 221–254.
- Sanders, Ted, Jentine Land & Gerber Mulder. 2007. Linguistic markers of coherence improve text comprehension in functional contexts. *Information Design Journal* 15(3). 219–235.

- Sandrelli, Annalisa. 2010. Corpus-Based Interpreting Studies and Interpreter Training: a Modest Proposal. Em Lew Zybatow (ed.), *Translationswissenschaft-Stand und Perspektiven, Innsbrucker Ringvorlesungen zur Translationswissenschaft VI*, (vol 12), 69–90. Peter Lang.
- Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. Em *Proceedings of the international conference on new methods in language processing*, 44–49.
- Schmidt, Thomas. 2004. Transcribing and annotating spoken language with EXMARALDA. Em *Proceedings of the LREC-Workshop on XML based richly annotated corpora*, ELRA.
- Setton, Robin. 1999. *Simultaneous Interpretation: A Cognitive-pragmatic Analysis*. John Benjamins.
- Shlesinger, Miriam. 1998. Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies. *Meta* 43(4). 486–493.
- Shlesinger, Miriam. 2009. Towards a definition of Interpretese: An intermodal, corpus-based study. Em Gyde Hansen, Andrew Chesterman & Heidrun Gerzymisch-Arbosgast (eds.), *Efforts and models in interpreting and translation research: A tribute to Daniel Gile*, 237–254. John Benjamins.
- Simões, Alberto & José João Almeida. 2007. Avaliação de alinhadores. Em Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*, 219–230. IST Press.
- Timarová, Šárka. 2005. Corpus linguistics methods in interpreting research: A case study. *The Interpreters's Newsletter* 13. 65–70.

CONTACTOS

Sílvia Araújo

Instituto de Letras e Ciências Humanas, Universidade do Minho
saraujo@ilch.uminho.pt

Ana Correia

Instituto de Letras e Ciências Humanas, Universidade do Minho
ana.moutinho@ilch.uminho.pt