

LINGUÍSTICA

REVISTA DE ESTUDOS LINGUÍSTICOS
DA UNIVERSIDADE DO PORTO
Vol. 19 (2024)

FICHA TÉCNICA

Linguística

Revista de Estudos Linguísticos da Universidade do Porto

Vol. 19, 2024

ISSN: 1646-6195

ISSN digital: 2182-9713

Periodicidade: Anual

Diretor:

António Leal

Secretariado editorial:

Vanessa Anachoreta

Editores:

Faculdade de Letras da Universidade do Porto

e Centro de Linguística da Universidade do Porto

Capa:

José Osswald

Os artigos publicados estão sujeitos a "peer review".

<https://ojs.lettras.up.pt/index.php/EL>

Esta publicação é financiada por fundos nacionais através da FCT - Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/00022/2020.

<https://doi.org/10.54499/UIDB/00022/2020>

ESPAÇO DA DIREÇÃO

O volume de 2024 da revista é composto de sete artigos, de temáticas e abordagens diversas, desde perspectivas históricas de fenómenos linguísticos até aos problemas encontrados nos (muito atuais) *Large Language Models*.

O primeiro artigo, "A gramaticalização de *a gente* em Português Europeu: um estudo baseado em corpus", é da autoria de Ana Filipa Fonseca e Clara Barros. Trata-se de um estudo em que se analisa o processo de gramaticalização da expressão "a gente" entre os séculos XIII e XX a partir de excertos retirados do *Corpus Informatizado do Português Medieval* e do *Corpus do Português*.

O artigo "The Impact of Corrective Feedback on EFL Learners' Acquisition of Regular and Irregular Past Simple Forms: Evidence from a Portuguese Context Using a Picture Description Task", da autoria de Ana Rita Faustino, apresenta um estudo que abordou os efeitos de três estratégias de correção de erros (correção explícita, reformulação e *prompts*) no ensino de tempos verbais em Inglês, por comparação com a ausência de correção. O objetivo desta investigação foi avaliar se essas estratégias promovem melhores resultados do que a ausência de *feedback*, se os efeitos variam entre formas verbais regulares e irregulares e por quanto tempo persistem os efeitos da metodologia usada.

Anette Kind e Simone Auf Der Maur são as autoras do artigo "Caso de estudo: Os docentes de alemão como língua estrangeira nas universidades portuguesas - impulsionadores do ensino digital em tempos de pandemia de Covid-19 e posteriormente", no qual analisam as dificuldades sentidas pelos docentes na digitalização do ensino de alemão como língua estrangeira em universidades portuguesas durante a pandemia de COVID-19, com base num estudo qualitativo sobre as experiências letivas à distância de professores e alunos.

O quarto artigo, "Os anos da mecanização (1940-1980): uma história da Tradução Automática", de Ronaldo de Oliveira Batista e Luciana Debonis, faz uma análise de como a tradução automática nasceu e evoluiu nas primeiras décadas da sua história.

Sueli Maria Coelho, no artigo “Preposição: uma categoria intermediária no continuum léxico > gramática”, propõe uma nova categorização para as 11 preposições essenciais mais produtivas do português do Brasil, baseada numa matriz de traços que combina três tipos de critérios, numa abordagem de tipo cognitivista.

Telmo Mória, em “Orações interrogativas com *qual*: análise gramatical e variação português europeu / português brasileiro”, analisa a distribuição do morfema interrogativo “qual” em interrogativas diretas e indiretas a partir de *corpora* jornalísticos portugueses e brasileiros. Para tal, o autor considera quatro contextos sintáticos principais, identificando algumas das regularidades subjacentes à utilização deste morfema em ambas as variedades.

Por fim, em “Analyzing GPT-4 Misinterpretations of Russian Grammatical Constructions”, Timofei Plotnikov relata uma experiência com o GPT-4, envolvendo juízos de gramaticalidade de expressões idiomáticas em russo. Os resultados mostram que, embora o modelo se comporte bem na generalidade dos casos testados, apresenta alguns erros, que são analisados ao longo do trabalho.

Este número inclui ainda dois textos, ambos da autoria de Isabel Margarida Duarte, lembrando dois eminentes linguistas que já não estão entre nós. O primeiro, em forma de recensão ao livro “...*Sedia la fremosa...*” *Textos de Teresa Brocado*, recorda a nossa colega da Universidade Nova de Lisboa, que demasiado cedo deixou a comunidade de linguistas portugueses. O segundo é uma breve nota sobre a vida e obra de Oswald Ducrot, figura maior na linguística do século XX.

A direção agradece a todos os colegas que se disponibilizaram a colaborar, com as suas avaliações, para a concretização de mais um número da revista *Linguística*.

SUMÁRIO

Espaço da Direção	03
Artigos	
A gramaticalização de <i>a gente</i> em Português Europeu: um estudo baseado em <i>corpus</i> <i>Ana Filipa Fonseca e Clara Barros</i>	07
The Impact of Corrective Feedback on EFL Learners' Acquisition of Regular and Irregular Past Simple Forms: Evidence from a Portuguese Context Using a Picture Description Task <i>Ana Rita Faustino</i>	34
Caso de estudo: Os docentes de alemão como língua estrangeira nas universidades portuguesas - impulsionadores do ensino digital em tempos de pandemia de Covid-19 e posteriormente <i>Anette Kind e Simone Madeleine Auf Der Maur</i>	60
Os anos da mecanização (1940-1980): uma história da Tradução Automática <i>Ronaldo de Oliveira Batista e Luciana Debonis</i>	85
Preposição: uma categoria intermediária no <i>continuum</i> léxico > gramática <i>Sueli Maria Coelho</i>	101
Orações interrogativas com <i>qual</i> : análise gramatical e variação português europeu / português brasileiro <i>Telmo Mória</i>	129
Analyzing GPT-4 Misinterpretations of Russian Grammatical Constructions <i>Timofei Plotnikov</i>	157
Recensões	
Teresa Brocardo <i>in memoriam</i> (1959-2023) <i>Isabel Margarida Duarte</i>	184
Notas	
Oswald Ducrot (27/11/1930 - 08/06/2024) <i>Isabel Margarida Duarte</i>	192

Artigos

A gramaticalização de *a gente* em Português Europeu: um estudo baseado em *corpus*¹

The Grammaticalization of 'a gente' in European Portuguese: A Corpus-Based Study

Ana Filipa Fonseca

Faculdade de Letras da Universidade do Porto (Portugal)

Centro de Linguística da Universidade do Porto

anafilipul@hotmail.com

Clara Barros

Faculdade de Letras da Universidade do Porto (Portugal)

Centro de Linguística da Universidade do Porto

mbarros@letras.up.pt

Recebido: 30/05/2024

Aceite: 15/07/2024

Publicado: 23/12/2024

ABSTRACT: This article intends to study the process of the grammaticalization of *a gente* which took place between the 13th and the 20th centuries (Lopes, 2004), comparing the different uses of the initial phase of grammaticalization, corresponding to the medieval period, and of the final phase (in the 19th and 20th centuries). To this end, a corpus was constituted composed of 156 excerpts extracted from the Computerized Corpus of Medieval Portuguese (CIPM) and the Corpus of Portuguese (CP) – subdivided into CP-Fiction and CP-News. In general, the results obtained corroborate conclusions already reached in other works on this evolutionary process (Nascimento, 1989; Menuzzi, 2000; Lopes, 2004; Costa & Pereira, 2013). However, we discovered some relevant aspects for this research: the use of a syntactic-semantic test to identify *a gente* with a semantic feature [Φ], the impact of regency on verbal agreement and the relationship between the uses and functioning of *a gente* within the various discursive traditions in which it arises. CP-News data confirm the characteristic tension of journalistic discourse between objectivity and approach to the reader, as the frequencies of the interpretations under analysis are very close (43,6.0% vs. 56,4%) and we even found a verb form in the first-person plural which agrees with *a gente*.

KEYWORDS: Grammaticalization; *A gente*; Verbal agreement; CIPM; Corpus do Português.

RESUMO: Neste trabalho, pretendemos estudar o processo de gramaticalização de *a gente*, que decorreu entre o século XIII e o século XX (Lopes, 2004), confrontando os usos distintos da fase inicial da gramaticalização, correspondente ao período medieval, e da fase

¹ Este trabalho foi financiado pela FCT - Fundação para a Ciência e a Tecnologia, I.P., projeto UIDB/00022/2020, e Bolsa de Doutoramento 2022.11010.BD de Ana Filipa Fonseca.

final (séculos XIX e XX). Para tal, constituímos um *corpus* composto por 156 excertos, extraídos do Corpus Informatizado do Português Medieval (CIPM) e do Corpus do Português (CP) - subdividido em CP-Ficção e CP-News. Em geral, os resultados obtidos corroboram o que é referido noutros trabalhos sobre este processo de evolução (Nascimento, 1989; Menuzzi, 2000; Lopes, 2004; Costa & Pereira, 2013). No entanto, descobrimos alguns aspetos relevantes para esta investigação: a utilização de um teste sintático-semântico para a identificação de *a gente* com traço semântico [Φ EU], o impacto da regência na concordância verbal e a relação entre os usos e funcionamentos de *a gente* com as várias tradições discursivas onde surge - os dados do CP- News confirmam a tensão característica do discurso jornalístico entre objetividade e aproximação ao leitor, pois as frequências das interpretações em análise são muito próximas (43,6% vs. 56,4%) e encontramos, inclusive, uma forma verbal na primeira pessoa do plural a concordar com *a gente*.

PALAVRAS-CHAVE: Gramaticalização; *A gente*; Concordância verbal; CIPM; Corpus do Português

Introdução

A classificação de *a gente* não é um assunto consensual na literatura, sendo discutidas diversas propostas (Sória, 2013, pp. 31-33): a expressão tem sido estudada como um nome (e.g. Nunes, 1919), um pronome indefinido (e.g. Melo, 1980), um pronome de tratamento (e.g. Bechara, 1967), forma substituta de *nós* (e.g. Cunha & Cintra, 1985), um pronome impessoal (e.g. gramática de Vázquez Cuesta e Mendes da Luz, de 1971) e, mais recentemente, como um pronome pessoal (e.g. Nascimento, 1989; Menuzzi, 2000; Pereira, 2003; Lopes, 2003, 2004).

A razão desta hesitação reside no longo processo de gramaticalização de *a gente*, ocorrido entre os séculos XIII e XX (Lopes, 2004). De acordo com Sória (2013, p. 33), podemos resumi-lo da seguinte forma: “→ De substantivo genérico (*gente*) → a pronome indefinido (*a gente*, como uma unidade gramaticalizada) → para, finalmente, pronome pessoal (*a gente*)”.²

Trata-se de uma expressão complexa que possui traços gramaticais e traços semântico-discursivos distintos, como é visível, entre outros aspetos, na concordância verbal (Menuzzi, 2000), onde se verificam três possibilidades de

² Neste estudo só analisaremos o processo de gramaticalização de pronome indefinido para pronome pessoal. A gramaticalização inicial do substantivo *gente* será considerada num projeto futuro, a par de outras estruturas impessoais, como *homem*, no Português Antigo, e *tu*, no Português Contemporâneo.

concordância: a terceira pessoa do singular, a primeira pessoa do plural e a terceira pessoa do plural (Costa & Pereira, 2013).

Assim, com este estudo, pretendemos confirmar o que é descrito na literatura relativamente ao processo de gramaticalização de *a gente*, confrontando duas sincronias onde *a gente* tem dois usos muito distintos: o período medieval, que, neste caso, compreenderá os séculos XIII a XVI, e o período contemporâneo, representado pelos séculos XIX e XX. Considerando sempre as propriedades enunciadas na literatura, aumentamos, contudo, os dados analisados no âmbito da gramaticalização de *a gente*, recorrendo a corpora diacrónicos; deste modo, observamos e analisamos outros aspetos potencialmente relevantes referentes ao uso e funcionamento de *a gente* nos referidos períodos.

O trabalho está organizado em duas secções centrais, a saber, o enquadramento teórico e o estudo propriamente dito, onde descrevemos o corpus e a metodologia, apresentamos a análise dos dados e os respetivos resultados e procedemos à sua interpretação. Teceremos ainda algumas considerações finais.

1. Estado da questão e enquadramento teórico

1.1. *A gente*: de pronome impessoal a pronome pessoal

A gente, utilizado para referir a primeira pessoa do plural ou conjuntos exclusivos de falantes, resulta do uso impessoal do nome *gente* com o artigo definido feminino singular *a*, tendo, com o tempo, perdido o significado lexical próprio do nome (e.o. Lopes, 1999; Posio, 2012).³ Na verdade, como sublinha Said Ali (1971, pp. 92-116), trata-se de um processo frequente alguns nomes transformarem-se em pronomes, a partir de contextos em que adquirem um sentido mais geral, sendo esse processo designado por gramaticalização (e.o. Hopper, 1991; Hopper & Traugott, 1993; Traugott, 2008).

³ Ver nota anterior.

Em Português Europeu (PE), esse processo foi demorado e gradual, tendo ocorrido entre os séculos XIII e XX (Lopes, 2004)⁴. De acordo com esta autora, no século XIII surgiu a primeira ocorrência de *a gente* pronominal – um caso único, num período onde existia também o nome *gente* no singular e no plural.⁵ No século XVI, à medida que a forma singular do nome se torna mais frequente do que a plural, já se encontram casos de interpretação ambígua, em que *a gente* já admite a leitura inclusiva (Lopes, 2004, p. 54). Só no século XIX é possível observar casos de *a gente* combinado com o possessivo *nossa* e respetivas variações, evidenciando, deste modo, traços de primeira pessoa do plural (Lopes, 2004, p. 62) – cf. o exemplo extraído dos dados da autora e transcrito em (1).

- (1) a gente vai mudar as nossas coisas para o terreno. (Séc. XX, Mendes 1981:118).⁶

Por fim, no século XX, *a gente* perde totalmente o traço formal de número – uma perda que se tinha vindo a manifestar gradualmente –, embora mantenha uma leitura semântica pluralizada (Lopes, 2004, p. 54).

No decorrer do processo de gramaticalização, os traços semântico-formais característicos do nome *gente* sofreram algumas alterações que descreveremos em seguida, com base em Lopes (2004).

Relativamente ao traço de número, a subespecificação do traço, própria do nome, deixou de existir, dado que, contrariamente ao nome *gente*, a forma pronominalizada ocorre apenas no singular (cf. *a gente* vs. *esta gente* / *estas gentes*). Contudo, trata-se de uma mudança gradual: a partir do século XVI, com o desaparecimento da expressão *homem* enquanto indefinido, observou-se uma aceleração da perda do traço de número plural, bem como a subespecificação de

⁴ No Português Brasileiro (PB) esse processo terá sido mais célere. *A gente* nesta variedade do Português já está mais gramaticalizado. Em contrapartida, o PE conserva os seus usos nominais ao lado dos pronominais, conforme descrevem Posio & Vilkuna (2013).

⁵ Neste estudo, na análise dos dados CIPM, encontramos, pelo menos, um outro exemplo, no texto das Cantigas de Escárnio e Maldizer.

⁶ Ressalvamos que a formatação de todos os exemplos que surgem neste trabalho é da nossa inteira responsabilidade, pese embora a maioria não ser da nossa autoria.

número, presente até ao século XIX; no século XX, o uso de *gente* apenas no singular tornou-se categórico (Lopes, 2004: 54).

Assim, dá-se uma simplificação (frequente nos processos de gramaticalização, como sublinham Heine & Reh (1984, p. 42)), na qual a distinção gramatical permanece semanticamente presente: a pluralidade é gramaticalmente irrelevante, mas não semanticamente.

Quanto ao género, ocorrem dois fenómenos: formalmente, observa-se a perda do traço, mas ao nível semântico-discursivo verifica-se uma subespecificação, que não consta das propriedades do nome *gente*. Apesar de lexicalmente o nome *gente* se caracterizar por alguma especificação - o seu traço formal de género, [+fem], denota especificação feminina -, a expressão não se restringia apenas à referência de um só sexo, sendo atribuído o traço [+genérico], e, semanticamente, o traço [ΦFEM], visto referir-se a um grupo de pessoas, cujo género semântico não era necessariamente especificado. Durante o processo de pronominalização, perde-se o traço formal [+fem], ou seja, *gente* / *a gente* deixa de ter um género formal; no entanto, semanticamente, observa-se uma subespecificação correspondente ao traço [α FEM] - *a gente* admite referência tanto a homens como a mulheres. De facto, é possível encontrarmos *a gente* com construções predicativas com adjetivos de género feminino (cf. (2)) e com construções predicativas com adjetivos de género masculino (cf. (3)), com referência mista ou exclusiva a homens.

(2) A gente está devastada.⁷

(3) A gente está devastado.

Em relação ao traço de pessoa, *a gente* mantém o traço formal [Φeu] que já caracterizava o nome *gente*, conforme atestam construções com formas verbais de terceira pessoa do singular - ver (4):

(4) A gente chega de férias domingo.

⁷ Os exemplos de (2) a (5) são de nossa autoria, baseados em usos orais frequentes do PE.

Ao nível da interpretação, porém, o traço semântico de pessoa [Φ EU] dá lugar ao traço [+EU] na forma pronominal, incluindo, assim, o falante na sua referência. Repare-se que, mesmo não sendo frequente no PE padrão ou entre falantes escolarizados, admite-se a concordância verbal de *a gente* com a primeira pessoa do plural e, inclusive, a partir do século XIX, encontra-se, segundo a autora, a forma pronominal *a gente* combinada com formas pronominais correlatas (os possessivos *nosso(s) / nossa(s)*), como é visível em (5).

(5) A gente vai buscar as nossas roupas para sair.

Em suma, como resume Lopes (2004), e seguindo os conceitos de Hopper (1991, pp. 22, 28 e 30) referentes ao processo de gramaticalização, constata-se, por um lado, a persistência de algumas propriedades (como a concordância verbal com a terceira pessoa do singular e o caráter coletivo e indeterminado do nome *gente*) e a decategorização, uma vez que a expressão sofreu alterações, tendo perdido as categorias flexionais de nome e adquirido propriedades tipicamente pronominais.

1.2. *A gente*: concordância verbal

Os traços gramaticais de *a gente* não coincidem com os seus traços semânticos, pois os primeiros são de terceira pessoa do singular, enquanto que os segundos são de primeira pessoa do plural.⁸ Esta divergência entre traços⁹ define a concordância de *a gente* dado que ambos os traços estão presentes e ativos em determinadas situações, segundo Menuzzi (2000), o que desencadeia várias possibilidades de concordância (Pereira, 2002).

Nascimento (1989, pp. 485-486) afirma que o uso normativo consiste no recurso à terceira pessoa do singular; já a concordância com a primeira pessoa do

⁸ Ver secção anterior ("1.1. *A gente*: de pronome impessoal a pronome pessoal").

⁹ A divergência de traços não está associada a uma idiosincrasia de *a gente*, na medida em que não se trata de um caso único. O pronome pessoal *vocês* também denota divergência de traços gramaticais e traços semântico-discursivos, como atestam os exemplos de Costa *et al.* (2001):

- a. Vocês, tu e a Maria, comeram as laranjas.
- b. *Vocês, ele e ela, comeram as laranjas.
- c. *Vocês, tu e a Maria, comestes as laranjas.

plural é considerada típica de registos menos cultos. Costa & Pereira (2013), além de exemplificarem estas duas possibilidades (cf. (6a) e (6b)), apontam uma terceira, própria do dialeto micalense, nos Açores: *a gente* a concordar verbalmente com a terceira pessoa do plural, que seria como ilustramos em (6c)¹⁰.

- (6) a. A gente vai à praia.
b. A gente vamos à praia.
c. A gente vão.

Estes autores concluíram que, estando as duas primeiras formas de concordância presentes num mesmo dialeto, escolher uma em detrimento da outra é meramente opcional. Na verdade, todas estas possibilidades tornam a temática da concordância verbal de *a gente* um assunto complexo. Inclusive, Costa *et al.* (2001, p. 653) salientam a “dificuldade dos falantes em emitirem juízos de gramaticalidade, bem como o facto de muitos falantes evitarem construções deste tipo, substituindo a gente por nós”.

1.3. A gente: um pronome diferente

Conforme referido na introdução, a classificação de *a gente* é motivo de discussão devido à divergência entre traços gramaticais e traços semântico-discursivos da expressão¹¹. Destacamos a classificação de pronome pessoal, mais recente e já mais ou menos estabilizada (ver, entre outros, Nascimento (1989), Menuzzi (2000), Pereira (2003) e Lopes (2003, 2004)), que iremos desenvolver nesta subsecção.

Nascimento (1989) apresenta alguns argumentos a favor do estatuto pronominal de *a gente*, tais como a possibilidade de *a gente* ocupar posições de sujeito, complemento direto, complemento indireto ou complemento preposicionado e a concordância verbal com formas de terceira pessoa do singular e/ou primeira pessoa do plural, que evidencia a tendência de *a gente* se fixar como

¹⁰ Este exemplo é de nossa autoria.

¹¹ Ver penúltimo parágrafo da Introdução, que se encontra na página 9.

forma de primeira pessoa do plural, coexistindo com *nós*. Refere igualmente a possibilidade de *a gente* anteceder o pronome reflexivo *se* e de designar uma “pluralidade mais ou menos definida”, arbitrária, a primeira pessoa do singular (*eu*) e a primeira pessoa do plural (*nós*).

Outros argumentos a favor do estatuto pronominal de *a gente* são a interpretação pronominal ou arbitrária e o facto de o género de *a gente* ser definido consoante a interpretação atribuída à expressão retomada (e.o. Menuzzi, 2000).

Contudo, alguns autores, como Taylor (2009)¹², mostram-se resistentes ao estatuto pronominal de *a gente*, já que a expressão apresenta, de igual modo, um comportamento atípico de pronome em duas situações: com quantificadores (ver exemplos (7) e (8) retirados de Costa & Pereira (2013, p. 167)) e com o que o autor designa “teste ‘we man’”. Este teste consiste na coocorrência de um pronome plural com um nome plural simples (“bare plural noun”), como *portugueses*. Apesar de, regra geral, a construção resultar gramatical com *nós* e *vocês* (ver (9)), tal não acontece com *a gente*, como ilustra (10), sendo novamente irrelevante a concordância verbal neste contexto.

- (7) a. Eles dois vão ao cinema.
b. Nós dois gostamos de pizza.
c. Vocês três viram a Maria.
- (8) a. *A gente dois vamos ao cinema.
b. *A gente dois vai ao cinema.
c. *A gente dois gostamos de pizza.
d. *A gente dois gosta de pizza.
- (9) a. Nós portugueses cantamos muito.¹³
b. Vocês estudantes não têm aulas.

¹² O autor defende que *a gente* é um DP que ora pode ter traços singulares (a.), ora pode ser apositivo de um pronome vazio de primeira pessoa do plural (b.). Ou seja, esta proposta justifica, por um lado, o comportamento não-pronominal e, por outro, os padrões de concordância verbal possíveis já aqui referidos.

a. [DP *a gente* [sg]]
b. [DP NÓSpl [DP *a gente* [sg]]]

¹³ Os exemplos (9) e (10) são igualmente retirados de Costa & Pereira (2013).

- (10) a. *A gente portugueses cantamos muito.
b. *A gente portugueses canta muito.

De facto, como podemos observar nos exemplos transcritos, apesar de ser gramatical o enunciado “Nós dois gostamos de pizza” (7b), resulta agramatical a construção “A gente dois gostamos de pizza” (8c) mesmo que a conjugação verbal seja na forma singular. De forma semelhante, embora seja aceitável “Nós portugueses cantamos muito” (9a), substituindo *nós* por *a gente*, a mesma construção resulta agramatical.

Costa & Pereira (2013), defendendo o estatuto pronominal de *a gente*, estabelecem um diálogo com o trabalho de Taylor (2009), de modo a justificar os resultados por ele obtidos. Para tal, os autores recorrem, por exemplo, à Teoria da Ligação (Chomsky, 1999) e concluem que é possível conciliar o estatuto pronominal de *a gente* com os dados observados, uma vez que as divergências no comportamento de *a gente* como pronome se devem a restrições de concordância, nomeadamente no caso de construções em que se observa falta de concordância em domínios locais.

1.4. Síntese

Em relação ao tema do processo de evolução de pronome impessoal a pronome pessoal, observa-se simultaneamente a conservação de algumas propriedades nominais (persistência) e a aquisição de propriedades tipicamente pronominais (de categorização).

Quanto ao tema da concordância verbal, fica comprovada a sua complexidade, dadas as possibilidades de concordância, justificadas pela ativação de traços semântico-formais distintos, e a consequente dificuldade dos próprios falantes em emitirem juízos de valor e optarem por uma delas.

Finalmente, no que diz respeito ao estatuto pronominal de *a gente*, verifica-se que, efetivamente, se trata de um pronome diferente, não devendo o seu estatuto ser colocado em causa. Na verdade, há uma explicação para os comportamentos atípicos observados em algumas construções: as restrições de concordância.

2. Estudo empírico

Nesta secção, apresentamos o estudo realizado. Assim, começaremos por descrever o corpus e a metodologia de análise; depois, apresentaremos a análise dos dados, seguindo-se a discussão dos resultados.

2.1. Corpus

Para a constituição do corpus tivemos em consideração algumas opções metodológicas, nomeadamente em relação ao âmbito temporal e aos contextos discursivos dos dados.¹⁴

No que diz respeito ao critério temporal, centramo-nos em dois momentos diferentes do processo de gramaticalização de *a gente* dados os objetivos deste trabalho, já mencionados na sua Introdução: a fase inicial, do século XIII ao século XVI¹⁵, e a fase final, que corresponde, principalmente, ao século XX, quando a gramaticalização da expressão já estará estabilizada.

Em relação aos contextos discursivos, optámos por abarcar várias tradições discursivas¹⁶ do mesmo período, a fim de perceber melhor o uso e funcionamento de *a gente* nos séculos referidos, pois uma linguística histórica que não considere a diversidade dos textos será, como refere Kabatek (2008, p. 9), “reductora y parcial”.

Neste sentido, escolhemos utilizar duas bases de dados distintas: o Corpus Informatizado do Português Medieval (CIPM) e o Corpus do Português (CP) - subdividido em CP-Ficção e CP-News.

O CIPM possui dados da época medieval de diversas tradições discursivas como cantigas, crónicas, textos jurídicos, textos apologéticos, entre outros. Já quanto aos dados extraídos do CP, os dois subcorpora permitem-nos analisar o uso

¹⁴ Dado que os fenómenos de gramaticalização surgem em contextos particulares de uso, é fundamental relacionar a evolução da forma estudada com um tipo de texto particular.

¹⁵ Consideramos o século XVI como fronteira temporal do período medieval, visto o período mais impactante para a gramaticalização de *a gente* compreender, na perspetiva de Lopes (2004), os séculos XVII, XVIII e XIX.

¹⁶ No âmbito deste trabalho, entendemos o termo *tradição discursiva* no sentido de Kabatek (2006, 2008). De acordo com o autor (Kabatek, 2006, p. 512), é a “repetição de um texto ou de uma forma textual ou de uma maneira particular de escrever ou falar que adquire valor de signo próprio (portanto é significável)”.

e funcionamento de *a gente* em duas tradições discursivas distintas, confrontando-as: a tradição discursiva jornalística, que é bastante específica, que intuímos evidenciar um uso acentuado de *a gente* impessoal, e a tradição discursiva ficcional, que procura reproduzir todos os usos da língua.

O corpus analisado neste estudo é composto por 156 excertos textuais, dos quais 70 (44,9%) foram extraídos do CIPM. Esse conjunto de casos representa os séculos XIII - com 46 exemplos (65,7%) -, XIV - com 1 único caso (1,4%) -, XV - também com 1 caso apenas -, XVI - ilustrado em 22 excertos (o que corresponde a 31,4% dos dados extraídos do CIPM).

Quanto aos excertos retirados da base de dados do CP: 58 (37,2%) são do subcorpus CP-Ficção, dos quais 1 (1,7%) data do século XIX e os restantes datam do século XX, e 39 (25,0%) são do subcorpus CP-News (todos eles datados do século XX).

Para a constituição do presente corpus, consideramos um único critério de seleção: recolher todos os segmentos, quer do CIPM, quer do CP, onde *a gente* funcione como um bloco. Conciliamos este critério com os seguintes critérios de exclusão: ocorrência de quantificadores a introduzir a expressão em análise; expressões resultantes da truncção de *a gente*. Ou seja, excluímos expressões como *a sua gente, toda a gente, a gent'*.

No que diz respeito à recolha e tratamento do corpus, extraímos os enunciados que iremos analisar de duas bases de dados, já referidas, o CIPM¹⁷ e o CP¹⁸, transcrevendo-os para um documento Excel. Posteriormente, com o objetivo de facilitar o manuseamento e, por conseguinte, a análise de dados, identificamos cada excerto recorrendo a uma codificação do tipo CIPMx (CIPM1, CIPM2, ...), para os dados extraídos do CIPM; CPFx (CPF1, CPF2, ...), para os dados extraídos do CP-Ficção; CPNx (CPN1, CPN2, ...), para os dados extraídos do CP-News.

¹⁷ Disponível em <https://cipm.fcsh.unl.pt/>.

¹⁸ Disponível em <https://www.corpusdoportugues.org/x.asp?c=2>.

2.2. Metodologia

Para a nossa análise, optámos por uma metodologia qualitativa e quantitativa, que nos permite descrever e verificar as interpretações de *a gente* (se é equivalente a *as pessoas*, *o povo*, *a humanidade*, *as pessoas* com o traço semântico de pessoa [ϕ EU], ou a *nós*, com o traço [+EU]) e a concordância verbal da expressão em estudo (se se recorre à terceira pessoa do singular, à primeira pessoa do plural ou à terceira pessoa do plural), de modo a constatar como se deu a gramaticalização de *a gente* em PE.

Quanto à anotação, esta foi realizada manualmente pelas próprias autoras.

2.3. Análise dos dados e apresentação dos resultados

Tendo em consideração os objetivos do presente estudo¹⁹, a análise dos dados teve por base dois critérios fundamentais, a interpretação e a concordância de *a gente*.

Deste modo, a análise subdividiu-se em dois momentos. Primeiramente, identificámos qual a leitura de *a gente*, tendo por base três valores possíveis: uma leitura em que *a gente* é semanticamente equivalente a *povo*, *população*, *humanidade* e *as pessoas*; uma leitura em que *a gente* é semanticamente equivalente a *nós*; uma leitura ambígua, em que não fosse possível identificar claramente a interpretação que *a gente* teria. Em seguida, procedemos à análise da concordância verbal, também ela baseada em três valores possíveis: *a gente* concorda com a terceira pessoa do singular; *a gente* concorda com a primeira pessoa do plural; *a gente* concorda com a terceira pessoa do plural.

A fim de facilitar a organização e a compreensão dos resultados obtidos, esta secção será subdividida de acordo com os dois parâmetros em análise já mencionados, a interpretação de *a gente* e a sua concordância.

¹⁹ Os objetivos deste trabalho podem ser encontrados na p. 9 deste trabalho.

2.3.1. Interpretação de *a gente*

Dada a heterogeneidade do corpus em análise, e para estruturarmos os dados de forma clara, esta subsecção será subdividida em três pontos, correspondentes, cada um deles, a um subcorpus: CIPM, CP-Ficção e CP-News.

2.3.1.1. Interpretação de *a gente* nos dados do CIPM

Dos 70 enunciados extraídos do CIPM, 69 (o que corresponde a 98,6%) representam *a gente* semanticamente equivalente a *povo*, *população*, *humanidade* e *as pessoas*, isto é, *a gente* com o traço semântico de pessoa [ϕ EU], tal como ilustrado em (11), (12), (13) e (14).

- (11) Pero que seja a gente d' outra lei [e] descreuda,
os que a Virgen mais aman, a esses ela ajuda. (CIPM1, século XIII)
- (12) E meteo todo este haver que assi andou
apanhando em milho, porque entom era refece, e pose-o em
guarda, e andou guarecendo pelos home~es boos de Galiza,
ataa que veo u~u ano mao em que valia muito o pam, e a
gente morria de fame. (CIPM58, século XIV)
- (13) El rey de Castela fez hyr o co~de estabre co~ toda a ge~te que levava
por assesegar seu reyno e a mostrar a parte que tiinha e~ Portugal. O
co~de estabre, co~ a gente que o iffante do~ Pedro ma~dava,
chegaro~ a Mayorga e foro~ onrradame~te recebydos del rey.
(CIPM66, século XV)
- (14) E acolhido na fortalleza, pomdo regra em seus mantimentos, foy
cercado por todas as partes d este rey dos de Dely, que jaa a este
tempo avya doze anos que lhe fazia a guerra, no quoyal cerco esteve
pouco tempo porque a gente que dentro na fortalleza estava, hera
muyta, e em pouco tempo gastara~o ho mantimento. (CIPM26, século
XVI)

No primeiro caso, a interpretação de *a gente* com um sentido equivalente a *povo* é acionada pelo modificador nominal introduzido pela preposição *de* (*a gente d'outra lei*), tal como ocorre em (11), e em (13) e (14) com as orações relativas introduzidas por *que* ("*a gente que o infante dom Pedro mandava*" e "*a gente que dentro na fortaleza estava*", respetivamente). Trata-se de contextos linguísticos semelhantes, na medida em que, quer o modificador, quer as orações relativas, restringem as entidades a que se referem dentro de um determinado universo referencial. Por outras palavras, referem-se a gentes específicas, caracterizadas por um conjunto de propriedades.

Já em (12), *a gente* consiste numa retoma anafórica da expressão "homens bons da Galiza", partilhando, assim, o seu referente.

Em relação ao enunciado no qual *a gente* possui o traço semântico de pessoa [+EU], sendo semanticamente equivalente ao pronome pessoal *nós*, este é do século XIII e encontra-se transcrito em (15).

(15) ((V15)) E cuidar d'el quen'o vir aqui,
que o vir andar assi calado,
ca nom sabe parte nem mandado
de tal justiça fazer qua[l] lh'eu vi:
leixou a gente adormecer entom
((V20)) e trasnoitou sobr'um hom'a Leon,
e fez sobr'el gram justiça logu'i. (CIPM13, século XIII)

Como se pode observar, é a presença do locutor, evidenciada na construção *lh'eu vi*, que implica a referida interpretação.

2.3.1.2. Interpretação de *a gente* nos dados do CP-Ficção

Dos 58 enunciados extraídos da base de dados do CP-Ficção, 39 (67,2%) - todos eles do século XX - representam *a gente* com o traço semântico de pessoa [+EU], como ilustrado em (16).

- (16) Guiomar bebia-lhe as palavras avinagradas como se elas fôsem de mel, Mas num segundo de clarividência feminina, interrompeu-o: - Porque veio, então? O médico é que se não deixou apanhar: - Nem sei. Dêstea gestos insensatos que a gente tem. Felizmente que a encontrei. Senão já me tinha ido embora. O instinto de Guiomar mandou-lhe que não insistisse. (CPF3, século XX)

Repare-se que, neste exemplo, *a gente* é semanticamente equivalente a um *nós* distinto, com sentido universal, que integra todas as pessoas, como demonstra a comutação de *a gente* por “*todos nós*”, que não altera o sentido do enunciado (cf. “Destes gestos insensatos que *a gente* tem.”²⁰ vs. “Destes gestos insensatos que *todos nós* temos.”).

Quanto a *a gente* com uma leitura semanticamente equivalente às expressões *povo*, *população*, *humanidade* e *as pessoas*, verificámos a ocorrência de 18 casos (o que corresponde a 31,0% dos dados extraídos do CP-Ficção) – cf. (17) e (18). Desses 18, apenas 1 (5,6%) data do século XIX (ver o exemplo (17)). Na verdade, esse é o único dado que conseguimos recolher referente ao século XIX.

- (17) Na Semana Santa j untava às rendas do chapéu a coroa amarga dos espinhos.. por serem estes, para a gente bem-nascida, dias de penitência e dor. E, diante de todo o livro ou de todo o quadro, sentia a emoção e formulava finamente o juízo, que no seu mundo, e nessa semana, fôsse elegante formular e sentir. (CPF23, século XIX)
- (18) E o que se lhe oferecia era Lisboa, o tumulto, os klaxons, os eléctricos, os ditos brutais dos catitas, os galanteios inconvictos e chulos, a gente apressada, o empurrão, o desdém das mulheres bem-vestidas, o futuro vazio, vazio. (CPF10, século XX)

²⁰ Optámos por traduzir o enunciado para português contemporâneo, de modo a facilitar a explicitação.

Note-se que, em ambos os casos, a *gente* surge modificado com um participípio passado (*bem-nascida* e *apressada*), que restringe um determinado conjunto de pessoas do qual o locutor se exclui.

Além disso, observámos 2 excertos textuais nos quais a *gente* tinha uma leitura particularmente ambígua. Veja-se (19) e (20):

- (19) É certo que o quadro não fazia qualquer sentido. Teria de pensar nisso. A arte não era brincadeira, não era coisa com que se brincasse. A gente olhava para pinturas e esculturas. O que aquilo significasse não teria qualquer significado, porque a presença do significado anularia o valor do quadro, isto é o que lhe pregara Arean, o especialista de Feraldo, o grandessíssimo ladrão. <> (CPF1, século XX)
- (20) - Olhe, pai - levantara-se e interrompeu-se para beijar a mãe que entrava com a terrina da sopa - boa noite, mãe, trazes uma sopa toda cheirosa. - E para o pai: - Eu já nem sei se o melhor não será aceitar a gente seja o que for. Embora tivesse dito " a gente " referia-se a Vasco e a Dulce, porque sobre si próprio tinha ideias assentes. (CPF9, século XX)

Em (19), a *gente* pode ser interpretado tanto com o traço semântico de pessoa [ϕ EU] como com o traço semântico [+EU], pois não há nenhuma pista linguística que force uma ou outra leitura.

Por oposição, em (20), apesar de, inicialmente, a *gente* ter como leitura preferencial o sentido semanticamente equivalente a *nós* - o contexto, uma conversa entre pais e filho, leva o leitor a interpretar a *gente* como expressão que se refere aos filhos -, o segmento seguinte esclarece as entidades a que o enunciador se estava a referir - "Embora tivesse dito " a gente", referia-se a Vasco e a Dulce, porque sobre si próprio tinha ideias assentes." -, explicitando que a *gente* se refere a um grupo de pessoas do qual o locutor se demarca.

2.3.1.3. Interpretação de *a gente* nos dados do CP-News

Os 39 enunciados extraídos da base de dados do CP-News datam do século XX. Destes 39, a maioria, mais concretamente 22 (o que corresponde a 56,4%), apresenta *a gente* com traço semântico de pessoa [+EU], como o exemplo transcrito em (21).

(21) Mas quando baixava a bandeira, ala para a trincheira que começava a chover morteirada. Andávamos todos a bater-las, eles eram tão obrigados como a gente. (CPN6, século XX)

Como se pode observar, é a forma verbal na primeira pessoa do plural (*andávamos*) que implica a referida interpretação de *a gente*.

Os restantes 17 excertos textuais (43,6%) evidenciam *a gente* com um sentido semelhante a *povo*, *população*, *humanidade* e *as pessoas*, isto é, com um traço semântico de pessoa [ϕEU], conforme ocorre em (22) e (23):

(22) Que o dissesse alto e bom som, para que a gente insatisfeita ficasse de uma vez por todas convencida de que agora as coisas não são assim. (CPN7, século XX)

(23) Interrogam-se também as pessoas sobre as razões que terão levado a gente da aldeia a cometer um crime de autoria colectiva. Há quem atribua as causas ao falhanço das instituições. (CPN31, século XX)

Em (22), tal como acontece em (17) e em (18), existe um modificador de nome (*insatisfeita*), que especifica as entidades a que *a gente* se refere, forçando a referida interpretação. O mesmo ocorre em (23), onde o modificador de nome é introduzido pela preposição *de* (*da aldeia*).

2.3.2. *Concordância verbal de a gente*

Conforme fizemos na subsecção 3.3.1., para apresentar os dados referentes à concordância verbal de *a gente* adotaremos uma subdivisão tripartida, consoante a fonte dos dados analisados: CIPM, CP-Ficção e CP-News.

2.3.2.1. Concordância verbal de *a gente* nos dados do CIPM

Dos 70 enunciados extraídos do CIPM, 53 (75,7%) evidenciam dados relativos à concordância verbal de *a gente*, sendo que não encontramos nenhuma informação referente à concordância verbal de *a gente* no século XV.

Em termos gerais, dos 53 enunciados mencionados, 41 (77,4%) apresentam formas verbais na terceira pessoa do singular, como em (24); 5 (9,4%) apresentam formas verbais na terceira pessoa do plural, como ilustrado em (25); e 7 (13,2%) evidenciam a coocorrência de ambas as formas verbais, terceira pessoa do singular e terceira pessoa do plural, como demonstrado em (26) - (32).

Nos 7 enunciados nos quais *a gente* coocorre simultaneamente com formas verbais na terceira pessoa do singular e na terceira pessoa do plural, verificámos uma tendência: independentemente da variação existente (por exemplo, na quantidade das formas verbais e na sua disposição sintática), nos 7 enunciados (o que corresponde a 100% dos casos) a(s) forma(s) verbal(ais) mais próxima(s) é(são) realizada(s) na terceira pessoa do singular e a(s) mais distante(s) na terceira pessoa do plural.

(24) E segund' as paravlas lle fez o son,
e depois cantó-o con gran devoçon;
e a tempestade quedou log' enton,
e perdeu en logo a gente temor. (CIPM50, século XIII)

(25) E a madr' enton chorando tan toste lle perdõou
e d' iren en romaria de grado llo outorgou;
mais dentro na ssa eigreja sol entrar nono leixou
a Virgen Santa Maria e ouv' enton a ficar

Non pod' ome pela Virgen tanta coita endurar ...
Fora, e entrou a madre. E a gente a veer-
o ve~eron e puxando o cuidaron y meter
per força; mais non poderon daquel logar o mover
que na eigreja entrasse por ben que fossen puxar.

Non pod' ome pela Virgen tanta coita endurar ... (CIPM65, século XIII)

- (26) E ela o rogo deles oyu e sa oraçon;
e estes que se querian mal, perdõaron-ss' enton,
e a gente que y era loaron de coraçõn
a Virgen de que Deus quiso naçer dia de Natal.

Sen muito ben que nos faze a Sennor esperital ... (CIPM16, século XIII)

- (27) E pois entrou en Tolosa, foi llogo fillar pousada
en casa dun grand' erege, non sabend' end' ele nada;
mas quando o viu a gente, foi ende maravillada
e disseron ao fillo: "Dest' albergue vos quitade."

Por dereito ten a Virgen, a Sennor de lealdade ... (CIPM21, século XIII)

- (28) Un trobador de Gasconna era, e trobava
al Con Symon e a muitos, si que sse queixava
a gente del, ca dizian que os de~ostava
mais quantos somos no mundo enquanto vivamos. (CIPM24, século
XIII)

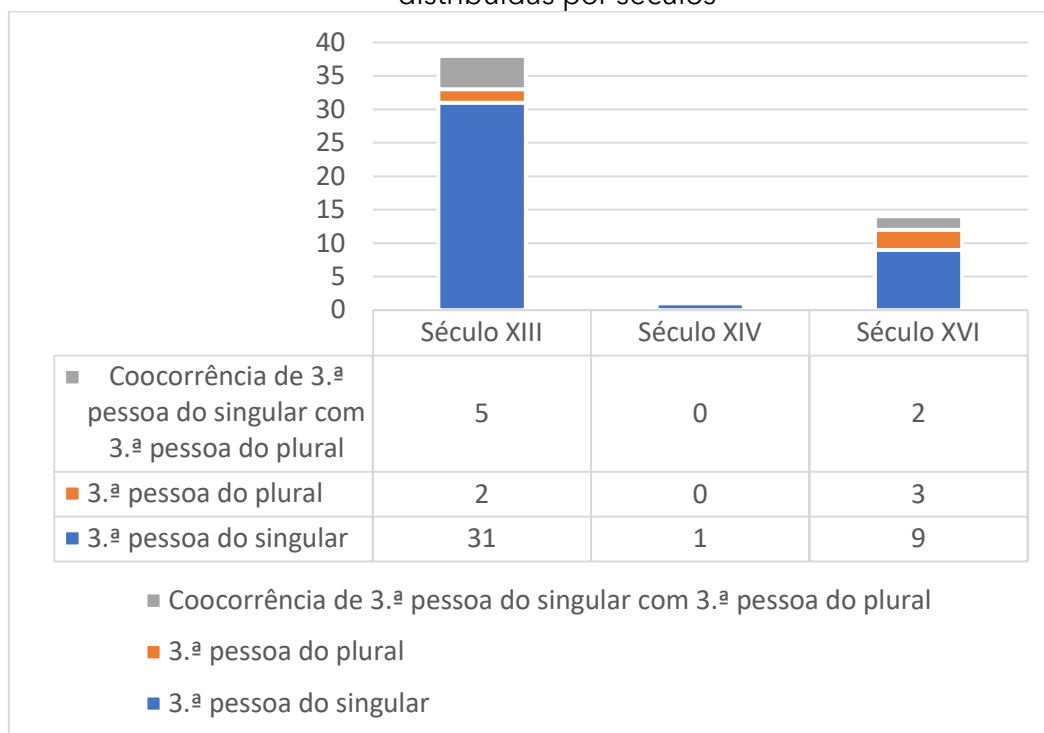
- (29) E acolhido na fortalleza, pomdo regra em seus mantimentos, foy
cercado por todas as partes d este rey dos de Dely, que jaa a este
tempo avya doze anos que lhe fazia a guerra, no quoad cerco esteve
pouco tempo porque a gente que dentro na fortalleza estava, hera
muyta, e em pouco tempo gastara~o ho mantimento. (CIPM26, século
XVI)

- (30) E cada sabbado sa~o obrigadas as molheres solteiras hir ao
paço a baillar e voltear diante do pagode d elrey, que tem dentro nas
suas casas, e a gente d esta terra jejua todos os sabbados, e na~o
comem todo dia, nem de noute, nem bebem augoa, se na~o comem
hu~u pouco de cravo, por amor do bafo (CIPM44, século XVI)

- (31) Tant' é grand' a sa mercee da Virgen e sa bondade ...
 Que aly u u catavan andou ele muit' agy~a
 tres vegadas a eigreja da Virgen Santa Rey~a
 a derredor; e a gente, que lle ben mentes ti~ia,
vírono como entrou dentro, mostrando grand' omildade. (CIPM61,
 século XIII)
- (32) E o colo con o braço tan forte se ll' estorceu,
 que en pees estar non pode e log' en terra caeu;
 mas a gente que viu esto o fillou e o ergueu,
 e correndo à eigreja o levaron de randon.
 Par Deus, muit', é gran dereito de prender gran[d]' ocajon ... (CIPM64,
 século XIII)

Relativamente à frequência de formas verbais encontradas nos dados do CIPM distribuídas por séculos, veja-se o Gráfico 1:

GRÁFICO 1 - Frequência de formas verbais encontradas nos dados do CIPM distribuídas por séculos



(Fonte: elaborado pelas autoras)

Repare-se que observámos apenas 1 concordância verbal (o que corresponde a 1,9% dos dados) que date do século XIV.

2.3.2.2. Concordância verbal de *a gente* nos dados do CP-Ficção

No que diz respeito aos dados do CP-Ficção, foi possível analisar, tendo por base a concordância verbal de *a gente*, 40 enunciados dos 58 que extraímos, o que corresponde a 69,0%. Esses exemplos datam do século XX, pelo que não encontramos nenhuma informação referente à concordância verbal de *a gente* no século XIX.

Nestes 40 enunciados, destaca-se a frequência de formas de terceira pessoa do singular (esta concordância verbal é evidenciada em 38 enunciados - 95,0% dos dados). A título de exemplo, veja-se o caso transcrito em (33).

Apenas 2 enunciados (5,0%) apresentam *a gente* com formas verbais na primeira pessoa do plural, como em (34).

De facto, não foi possível observar nenhum segmento textual onde *a gente* estabeleça uma relação de concordância com formas verbais na terceira pessoa do plural.

(33) A gente que velava o corpo de Hilário, no primeiro andar, foge precipitadamente. (CPF24, século XX)

(34) Chamaram, insistiram, gritaram por ele, nenhuma resposta: continuava ali, como que aparafusado ao leito do rio, certo apenas de que ninguém o havia de matar sem morrer também. - A gente somos malta da Calambata, Serrano. Baixa a arma e sai daí. Acabou a guerra. (CPF57, século XX)

2.3.2.3. Concordância verbal de *a gente* nos dados do CP-News

Já no que concerne aos 39 enunciados extraídos do CP-News, 24 deles (61,5%) denotam dados relativos à concordância verbal. Porém, tal como ocorre com os dados do CP-Ficção, também no CP-News não foi possível encontrar

nenhuma informação referente à concordância verbal de *a gente* no século XIX, pois os 24 enunciados que analisámos, tendo por base esse critério, datam do século XX.

Dos enunciados em questão, 23 (95,8%) apresentam *a gente* com formas verbais de terceira pessoa do singular, como demonstra (35).

O caso que resta, transcrito em (36), consiste na coocorrência de uma forma verbal de terceira pessoa do singular com uma forma verbal de primeira pessoa do plural. Note-se que, à semelhança do que verificámos nos dados extraídos do CIPM onde se observam coocorrências de formas verbais de pessoas diferentes, também neste enunciado a forma mais próxima é apresentada no singular e a mais afastada no plural.

(35) E é verdade que o filme sabe a pouco é como se não houvesse em ele tempo para a descida a os infernos de a corrupção, é como se não houvesse tempo para entrarmos em a cabeça de Henry Fonda e a gente bem queria Ah, como será diferente Corrupção! (CPN15, século XX)

(36) Quem não as marca, não falha. O Boavista merece tudo o que os jogadores têm feito pelo clube. Penso que os associados ficaram satisfeitos com a grande exibição, mas não ficaram satisfeitos com a finalização. É o trabalho da finalização, que a gente tem que afinar, para que não percamos tantas oportunidades», concluiu. (CPN30, século XX)

2.4. Discussão dos resultados

Nesta secção sumariaremos os principais resultados da nossa análise, estabelecendo um diálogo com os trabalhos já existentes citados no enquadramento teórico.

A emergência do uso dá-se no Português Antigo (século XIII), como evidencia o único exemplo encontrado no CIPM²¹. Isto prova que, efetivamente, o processo de gramaticalização de *a gente* se inicia muito cedo na língua, ainda que só no século XX a gramaticalização esteja concluída e a expressão *a gente* como pronome pessoal estabilizada.

Outro aspeto que indicia o processo de gramaticalização reside na evolução do traço semântico de pessoa, ao nível da interpretação, conforme descrito na literatura (Lopes, 2004), pois, enquanto os resultados obtidos nos enunciados que extraímos do CIPM revelam uma preferência preponderante pelo uso de *a gente* com traço semântico de pessoa [ϕ EU], os resultados obtidos nos dados extraídos do CP-Ficção denotam uma preferência pelo uso de *a gente* com o traço [+EU].²²

Em relação a interpretações ambíguas, não foi possível confirmar as afirmações de Lopes (2004), pois os exemplos datados no Período Clássico (século XVI) são todos de interpretação impessoal.

Verificámos também que poderá ser possível identificar a interpretação de *a gente* através de um teste sintático-semântico, pois, quando antecede modificadores oracionais ou nominais, *a gente* é semanticamente equivalente a *povo, população, humanidade* e *as pessoas* (como demonstram as expressões “*a gente d’outra lei*”²³, “*a gente que dentro na fortaleza estava*”²⁴, “*a gente bem-nascida*”²⁵, “*a gente apressada*”²⁶).

Em relação à concordância verbal de *a gente*, comprovámos a tripla possibilidade de concordância (e.o. Costa & Pereira, 2013), embora tenhamos constatado a existência de uma possibilidade *padrão*, particularmente frequente e transversal aos corpora analisados – a forma verbal na terceira pessoa do singular – que atesta o “uso normativo” mencionado por Nascimento (1989).

Além disso, a análise dos dados onde observámos coocorrência de *a gente* com formas verbais na terceira pessoa do singular e na terceira pessoa do plural,

²¹ Ver exemplo (15) na página 20.

²² Salientamos, uma vez mais, o objetivo comunicativo dos locutores do CP-Ficção de reproduzir o mais fielmente possível os diversos usos da língua no quotidiano e naquela época.

²³ Expressão extraída do exemplo (11).

²⁴ Expressão extraída do exemplo (14).

²⁵ Expressão extraída do exemplo (17).

²⁶ Expressão extraída do exemplo (18).

em simultâneo, induz-nos a questionar se a regência terá importância para a concordância verbal de *a gente* como tem para a sua concordância pronominal (ver Menuzzi, 2000).

Por fim, os resultados obtidos permitem-nos afirmar que o uso e funcionamento de *a gente* poderá relacionar-se com a tradição discursiva onde ocorre a expressão em causa, como exemplifica o CP-News. Nesta base de dados, as duas interpretações evidenciam um número aproximado de ocorrências ao nível da interpretação de *a gente*; e ao nível da concordância verbal, o facto de ser a única base de dados onde surge a concordância verbal na primeira pessoa do plural atesta a tensão própria do jornalismo entre objetividade e aproximação ao leitor.

Considerações finais

O objetivo deste trabalho consistia em estudar o processo de gramaticalização de *a gente*, confrontando os seus usos e funcionamento em dois momentos distintos deste processo – a fase inicial e a fase de estabilização da forma pronominal –, aumentar os dados analisados, recorrendo a outros corpora diacrónicos, e ainda apontar outros aspetos potencialmente relevantes que pudessem indicar novos caminhos a seguir na investigação.

De uma maneira geral, os resultados obtidos corroboram o que é referido noutros trabalhos que tratam a gramaticalização de *a gente* e o seu uso e funcionamento.

Fizemos, também, descobertas significativas para a investigação da gramaticalização de *a gente* em Português Europeu. Referimo-nos à aplicação de um teste sintático-semântico para identificar *a gente* com o traço semântico [ϕ EU] – evidenciado pela presença de modificadores –; à influência da regência na concordância verbal, de maneira semelhante ao que ocorre na concordância pronominal de *a gente* (Menuzzi, 2000); e finalmente à correlação entre os usos e funções de *a gente* nas diversas tradições discursivas em que aparece, tal como ilustrado pelo CP-News. De facto, as condições de produção e de circulação de um determinado discurso podem ter influência na interpretação e na concordância verbal de *a gente*. Os nossos resultados referentes especificamente ao CP-News

sugerem que as escolhas associadas a *a gente* evidenciam a tensão característica do discurso jornalístico de procurar objetividade ao mesmo tempo em que se busca uma relação de proximidade com o leitor, uma vez que as duas interpretações apresentam frequências mais próximas (43,6% e 56,4%). Um outro aspeto digno de nota é a ocorrência de uma forma verbal na primeira pessoa do plural, que reforça essa tentativa de se aproximar do leitor.

No decorrer deste trabalho, surgiram algumas pistas relevantes para aprofundar o estudo do uso e funcionamento de *a gente* no âmbito do seu processo de gramaticalização: além de ser necessário alargar o escopo da nossa análise a novos corpora para atestar as hipóteses descobertas já referidas, é relevante fazê-lo para estudar o aparecimento da ambiguidade e o designado “período intermédio” da gramaticalização, que antecipa a fase de estabilização; já num ponto de vista mais atual, questionamo-nos sobre qual será a situação contemporânea - terá vindo a diminuir o uso de *a gente* ou não? Ainda se manterá o destaque da concordância verbal com a terceira pessoa do singular ou a aproximação semântica da forma com o pronome pessoal *nós* gerou, com o tempo, mais ocorrências de concordância com a primeira pessoa do plural? Haverá uma variação entre estratos sociais e variedades diatópicas? E entre modalidades? Será a modalidade oral, frequentemente mais espontânea e familiar, mais propensa ao uso de *a gente* como sinónimo de *nós* e com a forma verbal na primeira pessoa do plural? Um estudo que tenha maior incidência na variação horizontal e vertical da língua poderá detetar mais casos e contribuir para esta investigação. Além disso, e dada a relação já mencionada entre as tradições discursivas e os fenómenos de gramaticalização, exemplificada, neste trabalho, pelo que observamos no CP-News, será igualmente relevante confrontar os usos e funcionamento de *a gente* em diferentes tradições discursivas, estabelecendo uma relação entre a sua frequência e a tipologia textual.

Em suma, reconhecemos que ainda há muito a fazer no que concerne à investigação deste tópico, uma vez que o processo de gramaticalização de *a gente*, como demonstrado neste artigo, foi um processo lento, gradual e bastante complexo, com implicações em múltiplos aspetos, entre os quais a concordância verbal. Deste modo, é importante considerar todas as novas direções e dimensões

que podem ser estudadas para enriquecer o conhecimento deste fenómeno da língua portuguesa.

Referências

- Bechara, E. (1967). *Moderna gramática portuguesa*. Nacional.
- Chomsky N. (1999). *O Programa Minimalista* (E. P. Raposo, Trad.). Caminho (Trabalho original publicado em 1995).
- Costa, J., & Pereira, S. (2013). *A gente: Pronominal Status and Agreement Revisited*. *The Linguistic Review*, 30(2), 161-184.
- Costa, J., Moura, D., Pereira, S., & Araújo, C. (2001). *Concordância com a gente: um problema para a teoria de verificação de traços*. *Actas do XVI Encontro Nacional da APL* (pp. 637-657). APL.
- Cunha, C. F., & Cintra, L. F. L. (1985). *Nova gramática do português contemporâneo*. Nova Fronteira.
- Heine, B., & Reh, M. (1984). *Grammaticalization and reanalysis in African languages*. Helmut Buske Verlag.
- Hopper, P. (1991). *On Some Principles of Grammaticalization*. In B. Heine, & E. Traugott (Eds.), *Approaches to Grammaticalization Volume I. Theoretical and methodological issues* (pp. 17-34). John Benjamins Publishing Company.
- Hopper, P., & Traugott, E. (1993). *Grammaticalization*. Cambridge University Press.
- Kabatek, J. (2006). *Tradições discursivas e mudança linguística*. In T. Lobo, I. Ribeiro, Z. Carneiro, & N. Almeida (Orgs.), *Para a história do português brasileiro. Volume VI: novos dados, novas análises. Tomo I*. EDUFBA.
- Kabatek, J. (ed.). (2008). *Sintaxis histórica del español y cambio lingüístico: Nuevas perspectivas desde las Tradiciones Discursivas*. Iberoamericana/Vervuert.
- Lopes, C. R. S. (1999). *A inserção de a gente no quadro nominal do português: percurso histórico* [Dissertação de doutoramento não publicada]. Universidade Federal do Rio de Janeiro.
- Lopes, C. R. (2003). *A inserção de a gente no quadro pronominal do português*. Iberoamericana/Vervuert.
- Lopes, C. R. S. (2004). *A gramaticalização de a gente em português em tempo real e de longa duração: retenção e mudança na especificação dos traços intrínsecos*. *Fórum Linguístico*, 4(1), 47-80.
- Melo, G. C. (1980). *Gramática fundamental da língua portuguesa*. Ao Livro Técnico.
- Menuzzi, S. M. (2000). *First Person Plural Anaphora in Brazilian Portuguese: Chains and Constraint Interaction in Binding*. In J. M. M. Costa (Ed.), *Portuguese Syntax: New Comparative Studies*. Oxford University Press.
- Nascimento, M. F. B. (1989). *A gente, um pronome da 4ª pessoa*. *Actas do Congresso sobre a Investigação e Ensino do Português* (pp. 480-490). Ministério da Educação / ICALP.
- Nunes, J. J. (1919). *Compêndio de Gramática Histórica Portuguesa (Fonética e Morfologia)* (8ª ed.). Livraria Clássica Editora.
- Pereira, S. (2002). *Concordância com a gente à luz da Morfologia Distribuída*. In M. A. P. Mendes, & T. Freiras (Coord.), *Atas do XVIII Encontro Nacional da*

- Associação Portuguesa de Linguística* (pp. 653-662). Universidade do Porto. APL.
- Pereira, S. (2003). *Gramática Comparada de a gente: variação no Português Europeu* [Dissertação de Mestrado não publicada]. Universidade de Lisboa.
- Posio, P. (2012). Who are 'we' in spoken Peninsular Spanish and European Portuguese? Expression and reference of first person plural subject pronouns. *Language Sciences*, 34(3), 339-360.
- Posio, P., & Vilkuna, M. (2013). Referential dimensions of human impersonals in dialectal European Portuguese and Finnish. *Linguistics*, 51(1), 177-229.
- Said Ali, M. (1971). *Gramática histórica da língua portuguesa*. Melhoramentos.
- Sória, M. (2013) *Nós, a gente e o sujeito nulo de primeira pessoa do plural* [Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa]. Repositório da Universidade de Lisboa.
https://repositorio.ul.pt/bitstream/10451/12218/1/ulfl157008_tm.pdf
- Taylor, M. (2009). On the pronominal status of Brazilian Portuguese *a gente*. *NYU Working Papers in Linguistics*, 2, 1-36.
- Traugott, E. (2008). Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In R. Eckardt, G. Jäger, & T. Veenstra (Eds.), *Variation, Selection, Development - Probing the Evolutionary Model of Language Change* (pp. 219-250). Mouton deGruyter.
- Vázquez Cuesta, P., & Luz, M. A. M. (1971). *Gramática da Língua Portuguesa* (A. M. Brito, & G. Matos, Trad.). Edições 70.

The Impact of Corrective Feedback on EFL Learners' Acquisition of Regular and Irregular Past Simple Forms: Evidence from a Portuguese Context Using a Picture Description Task

Ana Rita Faustino

Centro de Linguística da Universidade Nova de Lisboa (Portugal)

Instituto Superior de Contabilidade e Administração de Lisboa (Portugal)

Escola Superior de Educação de Lisboa (Portugal)

anaritafaustino@netcabo.pt

Recebido: 30/05/2024

Aceite: 28/07/2024

Publicado: 23/12/2024

ABSTRACT: Corrective Feedback (CF) is a common practice in the foreign language classroom and constitutes a topic of interest for both language teachers - who have to decide whether to correct their students, when and how - and for L2 researchers - who are interested in testing the efficacy of CF techniques in L2 acquisition. Conducted in a state school in Portugal, this quasi-experimental study investigated the effects of three CF strategies - recasts, prompts and explicit correction - and of no feedback on the acquisition of English regular and irregular past tense forms. 166 9th-grade Portuguese students (and their five teachers) took part in the study and were tested in a pretest-posttest design. Comparisons of group means using ANOVA revealed a significant effect for test time. Post-hoc analyses revealed that the Prompt group significantly outperformed both the Explicit Correction group and the Recast group in the production of regular past tense items and showed significant improvement between the Pretest and Posttest 2 in the use of irregular past tense.

KEYWORDS: English as a Foreign Language (EFL); Corrective Feedback (CF); Recast; Prompt; Explicit Correction.

Introduction

The present work focuses on the effects of Corrective Feedback (CF), "responses to learner utterances containing an error" (Ellis, 2006, p. 28), on the acquisition of regular and irregular English Past Simple forms. Research has shown strong support for the effectiveness of CF (Lyster, Saito & Sato, 2013; Pawlak, 2014), with its facilitative role being recognized by second language (L2) acquisition theories such as the Interaction Hypothesis (e.g. Long, 1996), the Output Hypothesis (e.g. Swain, 1995), Sociocultural Theory (e.g. Lantolf, 2006), and the Skill-Learning

Theory (e.g. DeKeyser, 1998). It has established itself as a key component in form-focused instruction, whether as a way of drawing the learner's attention to linguistic items in the context of activities whose focus is on communication, referred to by Long (1991) as *Focus on Form* (FoF), or as a means of reacting to the learner's erroneous utterances when developing accuracy in a more discrete-item based approach, which Long (1991) labelled *Focus on Forms*.

This study aims at investigating the effects of three feedback strategies as identified by Lyster & Ranta (1997), outlined below, and of no feedback on the acquisition of English past tense. The study further investigates whether these results persist over time. The CF strategies analyzed were as follows: *explicit correction*, in which the correct form is overtly supplied by the teacher, as illustrated in (1a); *recast*, when the teacher reformulates the learner's utterance incorporating the correction of the error, as shown in (1b); and *prompts*, a number of techniques that stimulate the learner to self-correct, including clarification request (1c), metalinguistic feedback (1d), elicitation (1e) and repetition (1f).

- (1) S: Last weekend I go to the cinema with my friends.
 - a. T: You should say "I went to the cinema with my friends".
 - b. T: Oh, you went to the cinema with your friends.
 - c. T: Pardon? Can you repeat?
 - d. T: What happens to the verb if you're talking about the past?
 - e. T: Last weekend, I...
 - f. T: I go?

These corrective turns may be followed by topic continuation by the same or another student, or by the teacher, or they may be followed by *uptake*, "a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the student's initial utterance" (Lyster & Ranta, 1997, p. 49). Uptake encompasses a wide array of learner responses to CF, from simple acknowledgements of a correction to reformulations of their errors. A type of uptake, *modified output*, consists in the learner's modification of a non-target-like form (Mackey, 2007). Lyster

& Ranta (1997) divided uptake into two categories: *repair*, when the error is correctly reformulated in a single turn; and *needs-repair*, when the uptake results in an utterance that still needs correction. Both categories are classifiable as modified output provided that the learner's turn entails modifications to the original error (Egi, 2010).

The target structure – the English past simple tense – remains a problematic one at intermediate and advanced levels, although it is introduced very early in the curriculum, and the distinction between regular and irregular forms enables us to study the effects of different CF strategies on a linguistic feature that is rule-based (regular forms) and item-based (irregular forms).

Taking part in this study were 166 intermediate level students of English as a Foreign Language (EFL) from a state school in Portugal, a context where there is a lack of research about oral CF. Each class received one of four different treatments – explicit correction, recast, prompt or no feedback – and was tested through a picture description task (PDT) in a pretest-posttest design.

The present article is organized as follows: Section 1 presents a brief overview of the research on the effectiveness of CF, focusing on laboratory and classroom studies; Section 2 presents the research questions and hypotheses; Section 3 provides information about the methodology, including the data collection tool, procedures for data collection and demographic information about the participants; and Section 4 presents the results of the statistical analysis for the PDT. The findings that emerged in the course of this study are presented in Section 5. The article ends with a summary of the main findings of the study and a brief discussion of pedagogical implications, limitations and future research directions.

1. Research on the effectiveness of CF

This section aims at providing a brief overview of empirical studies on CF. It starts with a short summary of selected laboratory studies, followed by a discussion of quasi-experimental studies that have contributed to the research on the effectiveness of CF.

1.1. Laboratory studies

There are a large number of laboratory studies which have yielded positive effects for recasts (e.g. Han, 2002; Ishida, 2004; Iwashita, 2003; Leeman, 2003; Long, Inagaki & Ortega, 1998; Mackey & Philp, 1998; McDonough & Mackey, 2006; Révész, 2007; Sagarra, 2007). Han (2002), for example, studied the impact of recasts on tense consistency in L2 output and found that recasts were successful at heightening the learners' (n=8) awareness and helped them in the correct use of past tense. The recast group outperformed the control group, which did not receive CF, following eight instruction sessions with intensive provision of recasts.

Focusing on three laboratory studies that compared recasts to other types of CF, Carroll & Swain (1993) compared the effects of explicit correction, recasts, prompts and no feedback on the acquisition of dative alternation in English (e.g., *John sent a letter to Mary* and *John sent Mary a letter*). All feedback groups outperformed the control group and the explicit correction group outperformed the recast and prompt groups in some measures. The recast and prompt groups had a similar performance. Lyster & Izquierdo (2009) investigated the effects of recasts and prompts on the acquisition of French grammatical gender. The analysis showed that the two types of feedback were equally effective. McDonough (2007) compared the effectiveness of recasts and clarification requests on the acquisition of English simple past verbs. The results demonstrated that both techniques were effective, with no significant difference.

1.2. Classroom studies

Doughty & Varela (1998) investigated the role of CF in the form of recasts as a type of FoF in two intermediate level ESL classes (n=34). The participants were divided into two groups: a FoF group (n=21) and a control group (n=13), which did not receive CF. Participants who received recasts outperformed those who did not receive CF in accuracy and total number of attempts at past time reference, particularly in the immediate posttest. Based on these results, the authors claim that providing CF in a timely manner that does not hinder natural communication

promoted learners' noticing of the target forms, which contributed to their higher performance.

Ellis *et al.* (2006) focused on the acquisition of the English regular past tense by three classes of lower-intermediate adult learners of English as a Second Language (ESL) in New Zealand (n=34). The classes were randomly divided into a recast group; a prompt group; and a control group, which continued with its regular activities. Overall, explicit CF was found to be more effective as it is more likely to be perceived as a correction move, thus generating "awareness of the gap between what was said and the target norm" (Ellis *et al.*, 2006, p. 329).

To investigate this relationship between noticing of CF and learning of past tense and question formation, four groups of high beginner college-level Francophone ESL learners (n=99) were assigned to a recast group, a prompt group, a mixed group (a combination of recasts and prompts) and a control group (no feedback) in a classroom quasi-experimental study by Kartchava (2012). The prompt and the mixed groups noticed more feedback than the recast group. As for the learning outcomes, accuracy levels increased more for past tense than for questions. Accuracy levels in the first posttest in the prompt and mixed groups improved for both features, while those in the recast group decreased both grammatical targets.

In a study involving 72 EFL Chinese learners, Yang & Lyster (2010) compared the effects of recasts and prompts on the acquisition of English regular and irregular past tense, using oral and written production measures. The prompt group, which received feedback in the form of metalinguistic clues, repetition, clarification request or elicitation, showed significant improvement in both the immediate and delayed posttests with respect to regular and irregular forms in both oral and written production. The recast group had significant short-term gains in the use of irregular past tense but no significant gains in the use of regular past tense in oral production. In written production, the recast group had significant short and long-term gains in the use of irregular past tense and short-term gains in the use of regular past tense. Given the low salience of regular past tense, the researchers hypothesize that recasts of irregular forms were more easily noticed by learners.

These studies suggest a positive effect for CF on promoting the noticing of the target feature (TF). Overall, explicit CF in the form of prompts seems to be more effective in generating learning outcomes than implicit feedback, such as recasts.

Additionally, several meta-analyses have revealed a medium to large effect of CF on acquisition, which was maintained over time. The benefits seem to be stronger in the foreign language (FL) than in the second language (SL) context (Li, 2010; Mackey & Goo, 2007) and more evident in the laboratory than in the classroom context (Li, 2010; Mackey & Goo, 2007). The results are not clear-cut as far as the differential effects of CF types are concerned: Norris & Ortega (2000) and Li (2010) found greater effects for explicit strategies - those that overtly indicated that the learner's output was not acceptable - than for implicit strategies - those that did not overtly draw the student's attention to his/her erroneous production. However, in Li's meta-analysis, while explicit strategies produced higher gains in immediate and short-delayed posttests, implicit ones had a greater effect on long-delayed posttests.

2. Research questions

To investigate the effects of CF strategies - explicit correction, recasts and prompts - and of no feedback on learners' acquisition of English regular and irregular past tense forms, the following research questions were formulated:

RQ1 Do the groups that received CF show higher gains in acquiring English past tense forms over the control group, which did not receive CF on past tense errors?

As described in Section 1, many classroom-based studies (as well as laboratory studies) have contributed to the perspective that CF assists L2 acquisition, as summarized in the meta-analyses on the topic (Li, 2010; Lyster & Saito, 2010; Lyster *et al.*, 2013; Mackey & Goo, 2007). Based on the results from previous research, we predicted that the feedback groups would perform better than the control group as far as the target structure is concerned.

RQ2 Do recasts, explicit correction and prompts have different effects on the acquisition of English regular versus irregular past tense forms?

Few studies have aimed at comparing CF effectiveness regarding regular and irregular past tense forms. In Section 1.2., some studies on the English past tense were reviewed. Whereas Ellis (2006) and Ellis *et al.* (2006) focused only on regular past tense, concluding that prompts are overall more effective than recasts, Yang & Lyster (2010) found that recasts promoted short-term gains for irregular forms but no significant gains for regular forms in oral production tasks and considerable short- and long-term gains for irregular forms and short-term gains for regular forms in written production tasks. Participants receiving prompts made significant gains on all measures and with respect to both regular and irregular forms. Considering the results of previous research, we predicted that prompts would have more significant effects than recasts on both regular and irregular forms and that recasts would be more effective with irregular forms than with regular forms.

3. Methodology

The present study, conducted in a state school in Portugal, sought to explore the impact of three CF strategies - explicit correction, recasts and prompts - and of no feedback on the acquisition of regular and irregular past simple forms, and to examine whether these results persist over time.

Seven classes of 9th-grade students (n=166) and their teachers (n=5) took part in the study. Student participants included 84 male students (51%) and 82 female students (49%). Their L1 was Portuguese, with most of the students reporting that they had been learning English for more than six years (76%) or from four to six years (24%), which indicates that the participants started learning the FL in primary school. All the classes followed the regular curriculum and, at the time of the study, had 135 minutes of English lessons per week, distributed between one 90-minute lesson and one 45-minute lesson. As a whole, the classes could be said to represent an intermediate level of proficiency in English, or B1, according to the Common

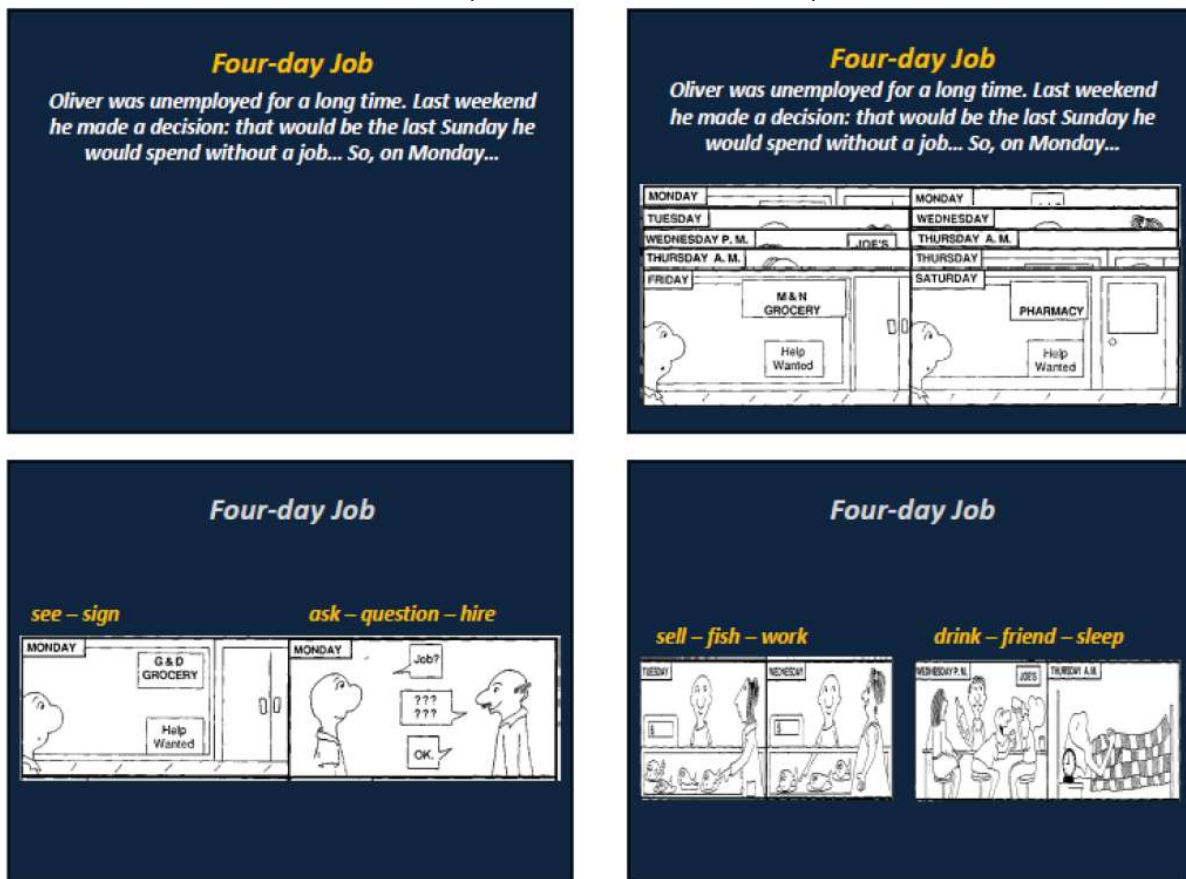
European Framework of Reference for Languages (Council of Europe, 2001), although they also comprised a few students who could be placed either above or below this threshold.

The five participating teachers were experienced EFL professionals who had taught English for eleven to thirty years, mainly in the state school context. To prepare for the implementation of the research study, teachers were assigned the CF strategy that they most naturally provided in class, which was ascertained through lesson observation by the researcher. Although the teachers were receptive to the idea of having the researcher attend their classes, it was not possible to implement video-recording or audio-recording in the present study. Given this restriction, several steps were taken to ensure that the instructional treatment was implemented as planned. The researcher met with each teacher beforehand to provide a more detailed explanation of the objectives of the study, the assigned feedback strategy and the data collection tools. The teachers received a document with a summary of the research aims and methodology and with specific guidelines on the provision of CF as far as their strategy was concerned. Moreover, the researcher was in regular contact with the participating teachers to discuss any questions they might have and to ensure the smooth implementation of the assigned CF strategy, having observed several lessons.

During the instructional period, the classes received the same amount of instruction - a total of nine hours over four weeks (one 90 minute-lesson and one 45-minute lesson per week). The instruction comprised activities included in the coursebook (Gonçalves, Coelho & Gonçalves, 2015) and some communicative activities to encourage the exposure to and the production of the target forms. Whenever a participant made an error in a target structure, the teachers provided CF by way of explicit correction (two intact classes), prompt (two intact classes) or recast (two intact classes). The control group (one intact class) did not receive any CF on past simple errors. It is worth noting that, while the CF was directed at individual learners, preference was given to activities that encouraged the whole class to be focused as much as possible on the speaker at moments of spoken production and interaction.

Students' performance on the accurate use of the past tense was measured in a pretest-posttest design with tests conducted before, immediately after and four months following the instructional intervention. For each testing session, a PDT was designed to elicit the production of sentences in the past tense. Each task included three cartoon stories with eight pictures each, depicting everyday situations, as illustrated in Figure 1. Learners had to tell the stories in the past using word cues: regular and irregular verbs in the infinitive, content words and adverbial phrases. There was no planning time after the pictures and word cues were presented, which potentially limited the learner's ability to draw on explicit knowledge of the TF (Ellis, 2005).

FIGURE 1- Excerpt from the Picture Description Task



(Source: adapted from Ligon & Tannenbaum, 1990)

4. Results

To investigate how the groups receiving different CF conditions performed in the PDT in the pretest, posttest 1 and posttest 2, their answers were computed and compared. A total of 17,388 occurrences were observed. Learners' responses were scored as either correct (1 point) or incorrect (0 points). In order to determine if there were any statistically significant differences among the pretest scores, a one-way analysis of variance (ANOVA)¹ compared the four groups. No significant differences were found among the groups on the PDT, $F(3, 157) = 1.008, p = .39$. Descriptive statistics for the participants' performance on the PDT with respect to regular and irregular verbs will be presented in the following subsections.

4.1. Regular Verbs

Descriptive statistics for the participants' performance on the PDT are presented in Table 1, which displays the mean, standard deviation and median as well as the lowest (minimum) and highest (maximum) scores of regular past tense accuracy for each group across the three testing times (pretest, posttest 1 and posttest 2). In the pretest the lowest mark was achieved by the Recast group ($n=54$), with a mean accuracy percentage score of 68.66% and the highest by the control group ($n=18$), with a mean accuracy percentage score of 80.77%. In posttest 1 the mean scores ranged from 72.28% to 85.73%, with the recast group being the lowest and the prompt group ($n=55$) the highest-scoring CF Type. In posttest 2 the mean scores ranged from 59.28% to 86.57%, with the explicit correction group ($n=34$) achieving the lowest mark and the prompt group the highest.

¹ According to the meta-analysis presented by Mackey & Goo (2007), ANOVA was the most frequently used statistical method for main analyses in CF studies.

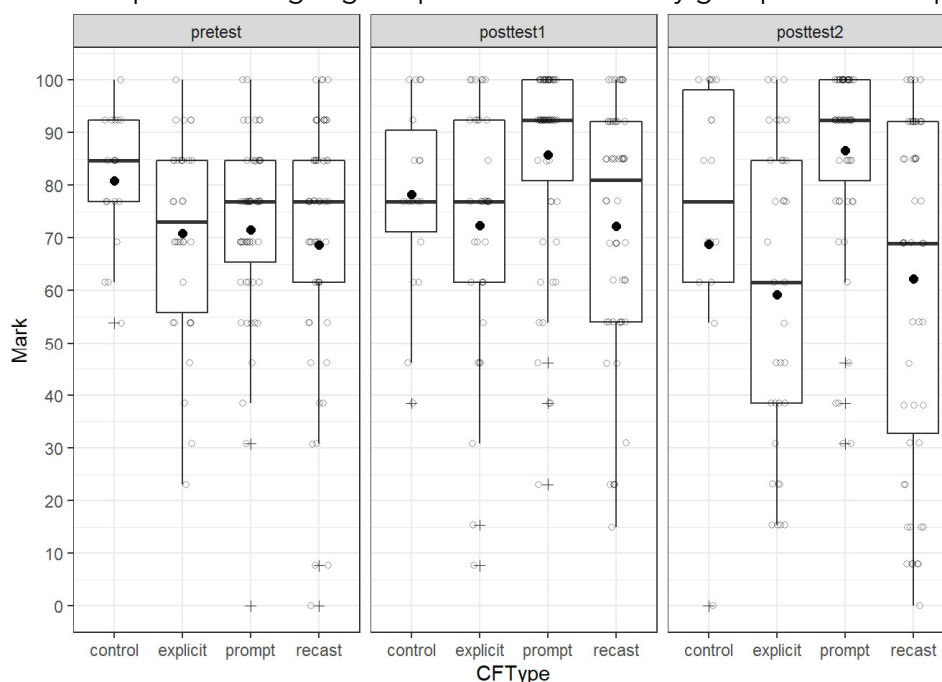
TABLE 1 - Mean percentage scores of regular past tense accuracy by group across test periods

Test	CF Type	Mean	Std. Dev.	Min.	Median	Max.
Pretest	Control	80.77	12.72	53.85	84.62	100
	Explicit	70.81	18.62	30.77	73.07	100
	Prompt	71.61	17.32	0	76.92	100
	Recast	68.66	25.53	0	76.92	100
Posttest 1	Control	78.21	17.94	38.46	76.92	100
	Explicit	72.40	23.96	7.69	76.92	100
	Prompt	85.73	18.87	23.08	92.31	100
	Recast	72.28	23.79	23.00	81.00	100
Posttest 2	Control	68.80	35.15	0	76.92	100
	Explicit	59.28	29.12	15.38	61.54	100
	Prompt	86.57	18.62	31.77	92.31	100
	Recast	62.20	32.76	0	69.00	100

(Source: created by the author)

Figure 2 graphically shows the mean, median, interquartile range, standard deviations and outlier values in the pretest, posttest 1 and posttest 2.

FIGURE 2 - Boxplot showing regular past tense scores by group across test periods

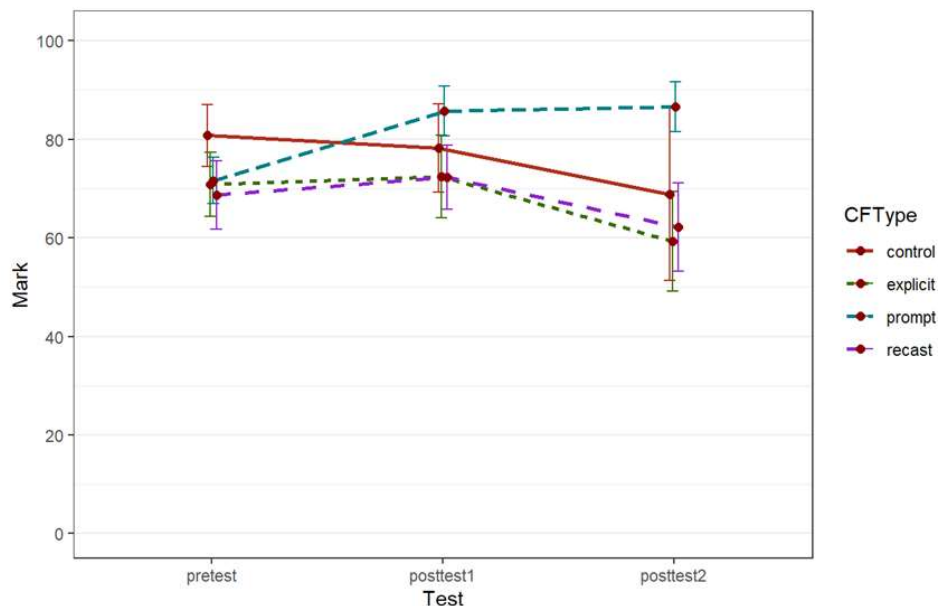


(Source: created by the author)

A two-way ANOVA with unbalanced design was performed, with the scores of the PDT as dependent variables. The statistical analysis was conducted on R. Test Time (pretest, posttest 1 and posttest 2) was a within-subjects independent variable, and CF Type (control, explicit correction, prompt, recast) was a between-subjects

independent variable. An alpha level of 0.05 was used. The results revealed a significant Test Time effect, $F(2, 157) = 4.108, p = .017$, a significant CF Type effect, $F(3, 157) = 11.594, p < .001$ and a significant interaction effect between Test Time and CF Type, $F(6, 157) = 3.256, p < .005$. Post-hoc analyses revealed that the prompt group significantly outperformed the explicit correction group ($p = .031$) and the recast group ($p < .005$) in the production of sentences containing regular past tense items. Figure 3 graphically displays the group accuracy percentage scores over time.

FIGURE 3 - Mean percentage scores of regular past tense accuracy by group across test periods



(Source: created by the author)

4.2. Irregular Verbs

Descriptive statistics for the participants' performance on the PDT are presented in Table 2, which displays means, standard deviations, median and the lowest (minimum) and highest (maximum) accuracy irregular past tense scores for each group across the three testing times (pretest, posttest 1 and posttest 2). In the pretest the lowest mark was achieved by the prompt group ($n=55$), with a mean accuracy percentage score of 54.86% and the highest by the control group ($n=18$), with a mean accuracy percentage score of 62.32%. In posttest 1 the mean scores

ranged from 56.04% to 73.04%, with the control group being the lowest and the prompt group the highest-scoring CF Type. In posttest 2 the mean scores ranged from 55.56% to 85.08%, again with the control group achieving the lowest mark and the prompt group the highest.

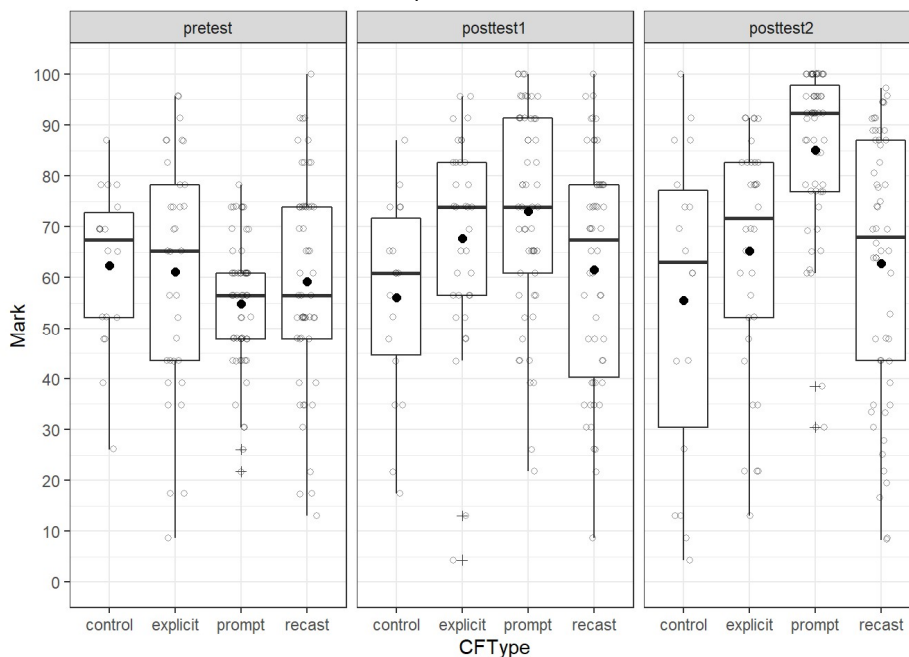
TABLE 2 - Mean percentage scores of irregular past tense accuracy by group across test periods

Test	CF Type	Mean	Std. Dev.	Min.	Median	Max.
Pretest	Control	62.32	15.92	26.09	67.39	86.96
	Explicit	61.13	23.57	8.70	65.22	95.65
	Prompt	54.86	12.95	21.74	56.52	78.26
	Recast	59.26	20.76	13.04	56.52	100
Posttest 1	Control	56.04	19.49	17.39	60.87	86.96
	Explicit	67.65	21.01	4.35	73.91	95.65
	Prompt	73.04	20.18	21.74	73.91	100
	Recast	61.59	22.69	8.70	67.39	100
Posttest 2	Control	55.56	31.02	4.35	63.04	100
	Explicit	65.22	23.25	13.04	71.74	91.30
	Prompt	85.08	17.29	30.43	92.31	100
	Recast	62.80	25.97	8.33	68.06	97.22

(Source: created by the author)

Figure 4 graphically shows the mean, median, interquartile range, standard deviations and outlier values in the pretest, posttest 1 and posttest 2.

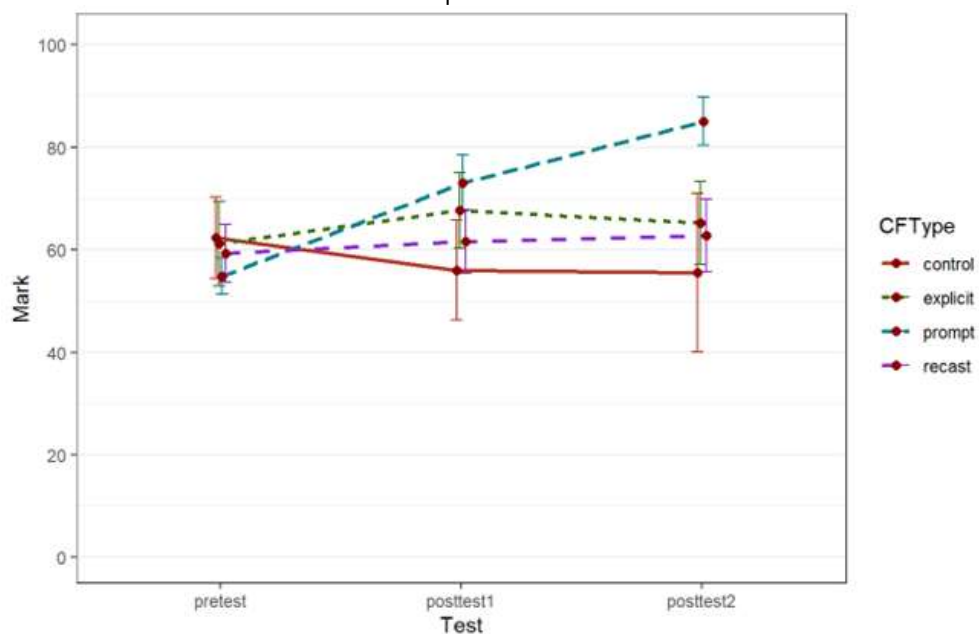
FIGURE 4 - Boxplot showing irregular past tense accuracy scores by group across test periods



(Source: created by the author)

A two-way ANOVA with unbalanced design was performed, with the scores of the PDT as dependent variables. R Studio was used to conduct the statistical analysis. Test Time (pretest, posttest 1 and posttest 2) was a within-subjects independent variable and CF Type (control, explicit correction, prompt, recast) was a between-subjects independent variable. An alpha level of 0.05 was used. The results revealed a significant Test Time effect $F(3, 157) = 12.547, p < .001$, a significant CF Type effect, $F(3, 157) = 8.129, p < .001$, and a significant interaction effect between Test Time and CF Type, $F(6, 157) = 5.886, p < .001$. Post-hoc analyses revealed significant gains in the prompt group between the Pretest and Posttest 2 ($p = .013$). Figure 5 graphically displays the accuracy percentage scores by group over time.

FIGURE 5 - Mean percentage scores of irregular past tense accuracy by group across test periods



(Source: created by the author)

5. Discussion

The present section is organized as follows: in subsection 5.1., the results of the different CF groups - control, explicit correction, prompt, recast - regarding regular and irregular past tense items are discussed. These results may be

influenced by a number of factors, such as the extent to which learners are familiarized with the TF, discussed in subsection 5.2., the variation in CF salience, analyzed in subsection 5.3., and the opportunities for modified output, discussed in subsection 5.4.

5.1. Regular and irregular items

The prompt group significantly outperformed the explicit correction group and the Recast group in the production of sentences containing regular past tense items in the PDT. Although the prompt group was the only group that managed to both improve its accuracy scores from the Pretest to Posttest 1 and maintain them from Posttest 1 to Posttest 2, it did not outperform the control group, which achieved relatively high scores from the Pretest onwards, maintaining its performance across the test times.

As far as irregular items are concerned, only the prompt group showed short- and long-term gains in the use of irregular past tense forms. These results are in line with those of Kartchava (2012), whose quasi-experimental study also included a picture description activity designed to elicit both regular and irregular past simple forms. In Kartchava's study, the accuracy levels increased in Posttest 1 for the prompt group and the mixed group, which received both prompts and recasts as CF treatment, whereas the scores for the recast group decreased. However, in the Results section the researcher did not differentiate between regular and irregular forms and no statistically significant differences were found between the groups.

Yang & Lyster (2010) also used a picture-cued narrative activity to test the production of regular and irregular past tense items. The researchers found that, regarding regular past tense, only the prompt group made significant gains, whereas all three groups significantly improved their irregular past tense scores over time. This testing instrument differed from the PDT in the present study, since the learners were provided with the complete narrative (including the past simple forms) as preparation and had three minutes to read the text before retelling the story based on word cues, which included all content words, adverbial phrases and verbs in their base form. There was, however, no planning time after the learners received

the word cues. The fact that more significant gains were found by Yang & Lyster (2010) might, to a certain extent, be explained by a memory effect. Although the researchers controlled for test-retest effect by using two versions of the test alternately over the three testing sessions, following the sequence ABA for all participants, each version contained exactly the same 9 regular and 18 irregular verbs. Having read the story three times, it is possible the learners may hypothetically have memorized some of the past simple forms, given that the retelling occurred immediately after the provision of the story. In the present study, learners were not provided with the complete narrative beforehand. The stories were presented to the students in the form of a PowerPoint presentation, with no planning time after the pictures and word cues were presented.

As discussed in Section 1.2., regular past tense forms are associated with low salience, low communicative value and high regularity, whereas irregular past tense forms have high salience, high frequency and low regularity (DeKeyser, 1998; Ellis, 2005). Therefore, we could expect recasts to be more effective in drawing attention to irregular than to regular forms, because irregular forms are more salient and it would potentially be easier to notice the difference between the student's own erroneous form and the correct form provided by the teacher. This was, however, not the case in the PDT administered in the present study. Prompts, on the other hand, were predicted to induce learners to notice both regular and irregular forms. The accuracy scores in the prompt group significantly increased for both regular and irregular forms in the PDT. In the case of explicit correction, the corrective intention is clearly communicated by the teacher and, therefore, these corrections could be expected to lead to more gains in accuracy. However, this was not the case in this task.

5.2. Familiarization with the target feature

Recasts are hypothesized to work better at providing exemplars of TFs that are relatively new to learners. The same would apply to explicit corrections, as they provide learners with positive evidence. Prompts, on the other hand, are predicted to promote a greater control over and faster access to already-acquired forms (Yang

& Lyster, 2010). During their experience as EFL learners, which, as previously established, for most of the participants in the present study began in primary school, the participants were exposed to classroom input containing exemplars of the TF and to the grammar rules involved. Therefore, like the participants in Ellis *et al.* (2006), Kartchava (2012) and Yang & Lyster (2010), the learners in the present study were at a stage in their learning process where they could take advantage of opportunities for TF production guided by CF.

5.3. Corrective Feedback salience

CF salience may also play a role in the results. As observed by Lyster & Ranta (1997), the implicitness of some recasts may contribute to these corrections being interpreted by students as repetitions of their correct answer. Lyster (2004, p. 404) also notes that “recasts of ill-formed utterances and repetitions of well-formed utterances together appear to confirm or disconfirm the meaning of a learner’s message”. The possibility exists that when the focus of the learners’ selective attention is on content, they may not perceive the corrective intention of recasts, leading them to ignore the linguistic evidence contained within them.

The results of this study differ from those of previous studies which found a positive effect for recasts. As discussed in Section 1.1., results that show the effectiveness of recasts stem especially from laboratory studies (e.g. Han, 2002; Mackey & Philp, 1998; Sagarra, 2007). In this research setting, CF is often provided consistently on a one-to-one interaction, which makes the corrective intention more evident to learners (Nicholas *et al.*, 2001). On the other hand, in the classroom setting, the implicitness or explicitness of recasts may be mediated by whether the teacher uses intonation, pauses for the learner to modify his or her output and also the content of the recast itself, i.e., if it is a short or a long recast and the number of changes it includes. For example, after analyzing the provision of recasts in four classroom settings, Sheen (2004) concluded that shorter recasts focusing on a specific word or phrase, recasts with an interrogative intonation and recasts that focused on phonological rather than grammatical or lexical errors were more likely

to lead to repair, with the results being corroborated by Loewen & Philp (2006) and Oliver & Grote (2010).

As illustrated in (2) below, in our observation of the classes in the recast group, we found several examples of relatively long recasts that involved multiple changes. In this interaction, the student's turn contains three errors. In the first utterance, the past simple of the verb *to be* (*was*) is missing and there is an omission error of the past simple form of the verb *want* (*wanted*). Moreover, the plural of the noun *child* is incorrectly formed (*children*). The teacher reacts by recasting the student's utterance, thus correcting the three errors and asking a follow-up question. These corrections occurred in a single turn, which may have hindered the learner's ability to process all the corrections and to notice the gap between his or her utterance and that of the teacher. Additionally, the teacher asks the student a question, for which the learner correctly provides an answer, which suggests that his or her attention was focused to a certain extent on reacting to the teacher's question rather than on acknowledging the difference between his or her turn and that of the teacher. In the student's response, we find no evidence that he or she has noticed a correction as the recast did not lead to uptake.

- (2) S: *When I a kid I want to be a teacher... To help childs learn new things every day.
T: You wanted to be a teacher when you were a kid? To help children learn new things every day... Are you not so sure anymore?
S: No... Teachers work a lot.

The several corrections in one teacher turn and the subsequent topic continuation by the teacher may pose a challenge to the learner in interpreting CF. Also, as argued by Goo (2012, p. 465), it is possible that responding to implicit CF moves such as recasts entails a greater cognitive effort, since "the noticing of recasts (...) necessitates cognitive control of attentional resources because it requires learners to engage in cognitive comparisons". Moreover, since students in the recast group did not receive clues on the nature and location of the error, even if they were

able to establish this cognitive comparison between their utterance and that of the teacher, they had to perform the task of identifying the error.

Furthermore, in other instances, a learner's well-formed utterance was repeated by the teacher not with a corrective intention, but with a communicative one: sustaining classroom interaction and maintaining its coherence. This is shown in (3).

(3) S: We went there together last week.

T: Oh, you went there together last week? Did you have fun?

S: Yes, a lot.

Thus, as far as recasts are concerned, the results of the present study differ from those of Doughty & Varela (1998), a quasi-experimental study described in Section 1.2. which investigated only this CF technique and found that the recast group revealed statistically significant gains in immediate and delayed posttests, whereas the control group showed no significant improvement over the same period. However, the "corrective recasting" in the treatment group consisted of two phases: first, repetitions of learners' imprecise output with rising intonation, and second, recasts to provide the target form. Thus, the recasting treatment provided leaned towards the explicit end of a continuum, since the erroneous form was stressed before another repetition of the learner's utterance incorporating the correction. The results were, nevertheless, interpreted as supporting evidence for the facilitative role of implicit forms of negative evidence. In the present study, recasts were operationalized as the teacher's reformulation, whether partial or full, of the learners' erroneous utterances that contained past tense items. Recasts included no repetition of ill-formed past tense and, thus, the CF provided was essentially implicit.

The corrective intention of prompts is generally easier to notice in the classroom setting (Ammar, 2008; Yang & Lyster, 2010), since they indicate that there is a problem with the learner's utterance but withhold the correct form, promoting noticing and modified output. This explicitness is a potential explanation for the greater effectiveness of prompts. Like the teachers in the recast and explicit

correction groups, the teacher in the prompt group was asked to provide immediate CF to the learners' past simple errors. However, the specific guidelines for the teacher in the prompt group were to provide the learners with clues that assisted them in reformulating their utterance. The teacher was also asked not to provide the correct form, unless the learner visibly struggled in self-correcting. These aspects added salience and explicitness to the prompt CF treatment. Firstly, they unequivocally pointed out the occurrence of an error. Secondly, once the learner was aware that his or her utterance needed repair, the aim of the teacher was to ensure that he or she was able to locate the error and self-correct. For example, the student in (4) incorrectly forms the interrogative by using both the auxiliary verb *did* and the past simple form *used*. The teacher in the prompt group treats the error through a prompt in the form of metalinguistic feedback, by reminding the student that a change in his or her utterance should be made due to the presence of an auxiliary verb in his or her question.

- (4) S: *What did you used to do at your grandparents'?
- T: Remember there is an auxiliary verb in the sentence, so...
- S: What did you use to do at your grandparents'?

As far as explicit correction is concerned, often there is no ambiguity regarding the teacher's intention of providing CF, as there is a clear indication that an error has occurred, as illustrated in (5). Nevertheless, the explicit correction group did not show any significant gains in accuracy. In (5), the student incorrectly uses the present simple form to refer to a past event. The teacher makes clear that the use of this form is incorrect and provides the student with the correct past simple form of the regular verb *start*.

- (5) S: *She starts to be an actress in 2008.
- T: It's not starts. It's started.
- S: She started.

5.4. Opportunities for modified output

Regarding the opportunities for modified output, recasts, explicit corrections and prompts differ in the extent to which learners attempt to self-correct. Despite being the most widely used CF type, the recast is the least likely to lead to uptake, followed by explicit correction. Prompts are usually more likely to elicit uptake from the student than the two input-providing strategies². Additionally, self-correction (or modified output) encourages the learner to test hypotheses when trying to reformulate the original utterance. Moreover, when attempting to produce an utterance that is more target-like, the learner may engage in deeper levels of processing than processing for meaning (Swain, 1995). Furthermore, practicing with a correct form after CF may promote automatization. Through reprocessing his or her original utterance, the learner's awareness of the TF is potentially increased.

In the case of recasts and explicit corrections, because the correct form is provided by the teacher, the repair that follows may consist mainly in an imitation of the teacher's utterance, without any deeper processing involved. In the interaction shown in (6), a student in the recast group incorrectly forms the past simple of the verb *take*, producing an error of regularization (**taked*). The teacher reacts with a full recast by repeating the student's utterance and incorporating the correct form *took*. Then, the student repeats the teacher's utterance in its entirety.

- (6) S: *I taked the bus to school yesterday at half past seven.
T: I took the bus to school yesterday at half past seven.
S: I took the bus to school yesterday at half past seven.

In this context, the students can have a different reaction if they repeat only the forms that were targeted in the CF episode, as shown in (7). In this case, the student in the recast group made a double marking error (**lefted*). The teacher

² In Lyster & Ranta's (1997) study, for example, only 31% of the recast episodes led to uptake; explicit corrections elicited uptake 50% of the time and prompts were similarly effective in leading to uptake: elicitation (100%), clarification request (88%), metalinguistic feedback (86%) and repetition (78%). These results are corroborated by Lyster (1998) and Sheen (2004).

reformulates the student's utterance to include the correct form *left*. Finally, the student reacts by repeating only the TF on which the teacher provided CF.

(7) S: *I lefted home at eight o'clock.

T: I left home at eight o'clock.

S: I left.

This type of modified output, which Gurzynski-Weiss & Baralt (2015) operationalize as partial modified output, i.e. when the learner repeats only the TFs that were targeted in the CF, may indicate that the learner has focused their attention on the error.

Conclusion

The present study was undertaken to contribute to the growing body of research on CF by investigating the effects of three CF types - Explicit Correction, Prompt and Recast - compared to no correction on the acquisition of English regular and irregular Past Simple forms. The findings suggest that prompts as a CF strategy can be effective in the FL classroom. These strategies can potentially be more effective than recasts when dealing with errors related to the past simple regular and irregular forms because the corrective intention of the teacher is more salient and because they provide opportunities for modified output. As previously discussed, several studies have shown recasts to be effective in the laboratory setting (e.g. Ishida, 2004; Iwashita, 2003; Leeman, 2003; Mackey & Philp, 1998; McDonough & Mackey, 2006) but less effective in the classroom context (e.g. Ammar & Spada, 2006; Ellis, 2007; Ellis *et al.*, 2006; Kartchava, 2012; Yang & Lyster, 2010). An implication that may be drawn is that recasts should be salient in terms of length and number of errors corrected. Additionally, teachers should be aware that a repetition of the recast by the learner does not necessarily mean that he or she has noticed the gap between their utterance and that of the teacher. Confirmation and/ or further explanation may be needed.

Furthermore, CF research is an area that may promote the dialogue between teachers and researchers. It is a relevant phenomenon for both, since instances of CF take place every day in most FL lessons and are an important tool used by language teachers in their interaction with students in an effort to maintain the flow of communication while also helping them improve their accuracy. Additionally, studies on CF inform L2 research on the importance of providing such correction for language development. Therefore, it is also in this interchange of ideas between researchers and teachers that relevant issues in the context of CF may arise, because “pedagogy should be research informed and [...] research will benefit from being pedagogy informed” (Ellis, 2017, p. 13).

The generalizability of the present research findings should be approached with caution due to several limitations. Firstly, the study involved a specific group of 9th-grade EFL students from a single state school in Portugal. Additionally, the assignment of classes to CF strategies was limited to two intact classes in the case of explicit correction, recast and prompts, and one intact class in the case of the control group. The absence of video or audio recordings of the lessons further constrained the ability to ensure consistent implementation of CF strategies. Participant fatigue may also pose limitations on the study's conclusions. Future studies might want to consider different populations and different instructional settings. CF is, to the best of our knowledge, an underexplored topic in the Portuguese context across all age groups and instructional settings. While the present study may have contributed to shedding some light on the effectiveness of CF strategies on a particular TF - the past simple in English - more research in FL classrooms in Portugal is clearly needed. Future research might investigate individual variables, other school years, geographical areas and FLs.

References

- Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research*, 12(2), 183-210.
- Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts and L2 learning. *Studies in Second Language Acquisition*, 28, 543-574.

- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: an empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, 15, 357-366.
- Council of Europe (CEFR) (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.
- DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge University Press.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition*. Cambridge University Press.
- Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: learner responses and language awareness. *The Modern Language Journal*, 94, 1-21.
- Gonçalves, M., Coelho, M., & Gonçalves, A. (2015). *iTeen 9*. Areal Editores.
- Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34, 445-474.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141-172.
- Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. *AILA Review*, 19, 18-41.
- Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 339-360). Oxford University Press.
- Ellis, R. (2017). Oral corrective feedback in L2 classrooms: what we know so far. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: research, theory, applications, implications* (pp. 3-18). Routledge.
- Ellis, R., Loewen, S., & Erlam, R. (2006) Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28, 339-368.
- Gurzynski-Weiss, L., & Baralt, M. (2015). Does type of modified output correspond to learner noticing of feedback? A closer look in face-to-face and computer-mediated task-based interaction. *Applied Psycholinguistics*, 36(6), 1393-1420.
- Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-572.
- Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form '-te i-(ru)' by learners of Japanese as a foreign language. *Language Learning*, 54, 311-394.
- Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition*, 25, 1-36.
- Kartchava, E. (2012). *Noticeability of corrective feedback, L2 development and learner beliefs* [PhD thesis, Université de Montreal]. Papyrus Institutional Repository. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/8531>
- Lantolf, J. (2006). Sociocultural theory and L2: state of the art. *Studies in Second Language Acquisition*, 28, 67-109.

- Leeman, J. (2003). Recasts and L2 development: beyond negative evidence. *Studies in Second Language Acquisition*, 25, 37-63.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60, 309-365.
- Ligon, F., & Tannenbaum, E. (1990). *Picture Stories: Language and Literacy Activities for Beginners*. Longman.
- Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness and effectiveness. *The Modern Language Journal*, 90, 536-556.
- Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). John Benjamins.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.
- Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, 82, 357-371.
- Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20, 51-81.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 19, 37-66.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 20, 37-66.
- Lyster, R., & Izquierdo, J. (2009). Prompts vs. recasts in dyadic interaction. *Language Learning*, 59, 453-498.
- Lyster, R., & Saito, K. (2010). Effects of oral feedback in SLA classroom research: A meta-analysis. *Studies in Second Language Acquisition*, 32, 265-302.
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46, 1-40.
- Mackey, A. (2007). Interaction as practice. In R. Dekeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 85-110). Cambridge University Press.
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford University Press.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, 82, 338-356.
- McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 323-338). Oxford University Press.
- McDonough, K., & Mackey, A. (2006) Responses to recasts: repetition, primed production, and language development. *Language Learning*, 56, 693-720.

- Nicholas, H., Lightbown, M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51, 719-758.
- Norris, J.M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
- Oliver, R., & Grote, E. (2010). The provision and uptake of different types of recasts in child and adult ESL learners. *Australian Review of Applied Linguistics*, 33(3), 26-41.
- Pawlak, M. (2014). *Error Correction in the Foreign Language Classroom: reconsidering the issues*. Springer Verlag.
- Révész, A. (2007). *Focus on form in task-based language teaching: recasts, task complexity, and L2 learning*. [Unpublished PhD dissertation]. Columbia University.
- Sagarra, N. (2007). From CALL to face-to-face interaction: the effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 229-248). Oxford University Press.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8(3), 263-300.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidhofer (eds.), *Principle and practice in applied linguistics: studies in honour of H. G. Widdowson* (pp. 125-144). Oxford University Press.
- Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, 32(2), 235-263.

Caso de estudo: Os docentes de alemão como língua estrangeira nas universidades portuguesas - impulsionadores do ensino digital em tempos de pandemia de Covid-19 e posteriormente¹

Case study: Teachers of German as a foreign language at Portuguese universities - drivers of digital education in times of the Covid-19 pandemic and beyond

Anette Kind

Faculdade de Letras da Universidade do Porto (Portugal)
Centro de Linguística da Universidade do Porto (Portugal)
anettekind@gmail.com

Simone Madeleine Auf Der Maur

Faculdade de Letras da Universidade do Porto (Portugal)
Centro de Linguística da Universidade do Porto (Portugal)
stome@letras.up.pt

Recebido: 04/06/2024
Aceite: 31/07/2024
Publicado: 23/12/2024

ABSTRACT: The Covid-19 pandemic has given a huge boost to the digitalization of university teaching. A central challenge was the *ad hoc* conversion and adaptation of face-to-face teaching scenarios to digital teaching formats, while respecting the guidelines of both the ministry and the universities. During this transition, lecturers took on a high level of responsibility, as they had to make autonomous didactic decisions in the implementation of digital teaching. This paper focuses on the role of teachers as the driving force behind the digitization of teaching German as a foreign language in Portuguese universities in the context of the pandemic. The data obtained through the qualitative study presented here allowed us to analyze the experiences of teachers and students of German as a foreign language during the pandemic. We conclude that, despite the rapid availability of training courses dedicated to the use of digital tools by universities and the commitment of teachers and students, digital teaching requires not only technical prerequisites and digital skills, but above all a digital didactic specifically orientated towards the teaching of foreign languages in the university context. In the near future, filling this gap will be a central challenge in the process of defining appropriate strategies for digital language teaching.

KEYWORDS: 'Digi-DaF' research project; Online teaching; Teaching German as a foreign language at university.

¹ Esta publicação é financiada por fundos nacionais através da FCT - Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/00022/2020.

RESUMO: A pandemia Covid-19 deu um enorme impulso à digitalização do ensino universitário. Um desafio central consistiu na conversão e adaptação *ad hoc* de cenários de ensino presencial para formatos de ensino digital, respeitando as diretrizes, tanto do ministério como das universidades. No decurso desta transição, os docentes assumiram um elevado nível de responsabilidade, na medida em que tiveram de tomar decisões didáticas autónomas na implementação do ensino digital. Este artigo centra-se no papel dos docentes como força motriz da digitalização do ensino de alemão como língua estrangeira nas universidades portuguesas, no contexto da pandemia. Os dados obtidos através do estudo qualitativo aqui apresentado permitiram a análise das experiências dos docentes e estudantes de alemão como língua estrangeira durante a pandemia. Concluímos que, apesar da rápida disponibilização de formações dedicadas à utilização de ferramentas digitais, por parte das universidades, e do empenho de docentes e estudantes, o ensino digital exige não só pré-requisitos técnicos e competências digitais, mas sobretudo uma didática digital especificamente orientada para o ensino de línguas estrangeiras no contexto universitário. A colmatação desta lacuna constitui, num futuro próximo, um desafio central no processo da definição de estratégias adequadas para o ensino digital das línguas.

PALAVRAS-CHAVE: Projeto de investigação "Digi-DaF"; Ensino *online*; Ensino de alemão como língua estrangeira nas universidades portuguesas; Competência digital; Didática do ensino digital.

Introdução

A pandemia Covid-19 colocou docentes e estudantes,² a nível mundial, e praticamente de um dia para o outro, perante desafios inesperados. Em Portugal, com o confinamento implementado a partir de março de 2020, o ensino em geral, e assim também o ensino universitário, passaram a funcionar exclusivamente *online*. Rapidamente, as universidades disponibilizaram formações dedicadas ao trabalho com ferramentas e plataformas digitais como Zoom ou Moodle, para os docentes poderem continuar a lecionar as suas aulas a distância.³ As dificuldades que daí

² Para simplificação da leitura utilizamos, neste artigo, o masculino genérico.

³ Moreira & Schlemmer (2020) definem o termo "ensino a distância", como caracterizado pela separação física e, por vezes, temporal, entre alunos e professores (p.10), baseando-se na definição de Moore & Kearsley (2011). Aplicam os mesmos critérios para a definição do ensino remoto (p. 8). Seguindo esta proposta, neste artigo, empregamos ambos os termos de forma sinónima. Usamos o termo "aulas online" para designar aulas síncronas via Zoom que foram implementadas como medida de emergência na pandemia Covid-19, uma vez que, no nosso estudo, apenas foi utilizada essa mesma plataforma.

Quando falamos de aulas/modelos híbridos, estamos-nos a referir aos modelos praticados nas fases com confinamento menos restritivo em que já se voltou, parcialmente, ao ensino presencial, respeitando medidas de segurança como maior distância física. Nas oito universidades inquiridas praticaram-se modelos híbridos muito diversos, sendo o mais complicado, na perspetiva dos estudantes, o modelo em que tiveram que assistir, no mesmo dia, a aulas presenciais, com menos

resultaram, tanto para docentes como para estudantes, eram particularmente acentuadas no ensino das línguas estrangeiras, que pressupõe uma substancial participação oral dos estudantes e a permanente interação entre os mesmos e os seus docentes.

Estas dificuldades, assim como o interesse em saber que estratégias os outros docentes de alemão como língua estrangeira desenvolveram nesta situação, motivaram seis docentes de alemão de quatro universidades portuguesas a iniciar um projeto de investigação, realizando um levantamento de dados que se prendem com a perspetiva de docentes e estudantes perante a nova situação de ensino-aprendizagem.⁴

1. Objetivos do projeto “Digi-DaF”

O projeto de investigação, financiado pelo Centro de Linguística da Universidade do Porto, intitulava-se, inicialmente, “Digi-DaF: Desafios do ensino universitário digital de línguas, pelo exemplo do ensino do alemão como língua estrangeira, durante a pandemia Covid-19 e posteriormente: boas práticas e modelos sustentáveis” e visava analisar as seguintes questões:

- Reações das universidades face às diretivas do Ministério de Educação nas fases de confinamento;
- Diferentes modelos de ensino e formatos de aulas aplicadas pelas diversas universidades (sobretudo nos modelos híbridos nas fases do confinamento menos restritivo);

participantes, e aulas online, sem que houvesse salas destinadas a essa finalidade, o que fez com que, caso fizessem parte dos estudantes destinados a assistir remotamente, ficassem sentados nos corredores da faculdade para assistir às aulas, sem possibilidade de participar oralmente, já que não havia nenhum isolamento acústico. Na perspetiva dos docentes, por sua vez, revelou-se mais complicado o modelo híbrido em que tiveram de integrar tanto estudantes presentes na sala como estudantes presentes por Zoom, o que, nas aulas de língua, não permitiu a interação entre estudantes dos dois grupos. Houve ainda modelos com aulas presenciais, apenas em certos dias da semana, recorrendo a salas maiores da universidade, enquanto o outro grupo executava tarefas atribuídas previamente, sem assistir às aulas.

⁴ Anette Kind, direção do projeto (Universidade do Porto), Ingrid Scholz (Universidade do Porto até 2017, Ludwig-Maximilians-Universität München), Simone Auf der Maur (Universidade do Porto), Katrin Herget (Universidade de Aveiro), Andrea Riedel (Universidade de Lisboa), Claudia Ascher (Universidade de Coimbra).

- Alterações e estratégias concretas introduzidas nas aulas online pelos docentes de alemão;
- Identificação de desafios e obstáculos com que tanto docentes como estudantes se depararam;
- Levantamento de modelos e estratégias que se revelaram proveitosos, na ótica dos docentes, e que foram bem recebidas pelos estudantes, com potencialidades de constituir um instrumentário para aulas de línguas estrangeiras após a pandemia (*best practices*).⁵

Pretendemos recolher dados tão representativos quanto possível, na perspetiva de docentes e estudantes, das oito universidades portuguesas públicas em que o alemão é disciplina nuclear de alguns cursos de licenciatura,⁶ que permitissem tirar conclusões sobre as condições em que se processou o ensino *online*.

Após análise dos dados do primeiro grupo inquirido, constatamos que, embora os docentes tenham posto em prática várias abordagens inovadoras para responder à nova situação do ensino remoto, prevaleciam as dúvidas e dificuldades que identificaram. Por isso, incluímos como um dos objetivos, formular os *desiderata* para identificar as condições imprescindíveis para a realização de aulas a distância em cenários futuros de ensino online ou híbrido, que poderão resultar tanto duma nova crise e/ou pandemia ou de opções deliberadas, motivadas pelas mais variadas razões, por exemplo, para aumentar a componente digital do ensino. Por isso, complementamos a designação do projeto, intitulado-se, doravante: “Digi-DaF: Desafios do ensino universitário digital de línguas, pelo exemplo do ensino do alemão como língua estrangeira, durante a pandemia Covid-19 e posteriormente:

⁵Os objetivos do projeto de investigação e a metodologia aplicada foram apresentados, de forma mais exaustiva, na comunicação “Projeto de Investigação inter-universitário Digi-DaF” (Parte 1): Desafios e Oportunidades de novos formatos digitais no Ensino-Aprendizagem do Alemão na Perspetiva dos Docentes” no âmbito do Congresso *Aprolínguas 2021*. O respetivo artigo, assim como os demais artigos resultantes deste congresso, entregues para publicação em junho de 2022, ainda aguardam publicação.

⁶Universidade do Porto, Universidade de Aveiro, Universidade de Coimbra, Universidade de Lisboa, Universidade do Minho (Braga), Universidade Nova de Lisboa, Universidade da Madeira e Universidade Católica Portuguesa (Lisboa).

boas práticas e *desiderata*". Até à data, já foram apresentados resultados parciais em vários congressos e simpósios.⁷

No presente artigo focamo-nos na formação disponibilizada pelas universidades (*workshops* e tutoriais) para assegurar a repentina passagem do ensino presencial ao remoto, bem como nos desafios com que os docentes universitários de alemão se viram confrontados, na busca de soluções adequadas, e no seu esforço para encontrar estratégias e procedimentos metódicos adaptados ao ensino digital. A análise destes aspetos permitirá tirar conclusões quanto ao grau de preparação pedagógico-didática dos docentes, perante os desafios do ensino remoto.

Por razões óbvias, não podemos nem pretendemos dar um panorama geral sobre todas as alterações resultantes da passagem do ensino presencial ao ensino remoto introduzidas a partir de março 2020, desde o ensino escolar ao superior e em todas as áreas científicas. No entanto, para melhor contextualização, vamos referir alguns estudos dedicados ao ensino remoto (emergencial) em geral e à didática que este mesmo envolve.

⁷ Kind, A., 27.10.2021: "Projeto interuniversitário Digi-DaF: Novos desafios para o ensino digital do Alemão"; Comunicação no Ciclo de conferências: "O Ensino de Línguas em Tempo de Pandemia: Transição, Transformação, novas Perspetivas", organizado pelo Centro de Línguas, Literaturas e Culturas, do Departamento de Línguas e Culturas da Universidade de Aveiro.

Kind, A, Herget, K, Scholz, I.: Projeto de Investigação inter-universitário „Digi-DaF“ (Parte 1): Desafios e Oportunidades de novos formatos digitais no Ensino-Aprendizagem do Alemão na Perspetiva dos Docentes.

Auf der Maur, S, Riedel, A. , Ascher, C.: Interuniversitäres Forschungsprojekt "Digi-DaF" (Teil 2): Modelle und Methoden des Fremdsprachenunterrichts in Zeiten von Covid-19 - Herausforderungen und Chancen". Comunicações no III *International Conference Aprolínguas 2021: "Foreign Language Teaching and Learning in the Classroom and online: Tradition, Perspectives and Challenges"*, 2-3.12.2021.

Kind, A., "Alemão como Língua Estrangeira nas universidades portuguesas: Interação entre continuidade, transformações e novos desafios", Conferência plenária no Congresso *SILAPP (I. Simpósio Internacional de Língua Alemã nos Países de Língua Oficial Portuguesa): "Contribuições do ensino e da aprendizagem da língua alemã na CPLP"*; na Universidade de São Tomé e Príncipe, 14-16/6/2023.

Auf der Maur, S., Kind, A.: "Os docentes de alemão como língua estrangeira nas universidades portuguesas - impulsionadores do ensino digital em tempos de pandemia de Covid-19 e depois", comunicação no *Congresso Bienal do CLUP (Centro de Linguística da Universidade do Porto)*, 2-4/11/2023.

2. Didática e pedagogia no ensino remoto desde o início da pandemia de COVID-19

Embora as abordagens de ensino a distância (e *blended-learning*) não sejam novas, ganharam uma relevância acrescida com as fases de confinamento durante a pandemia. Desde 2020, a nível mundial, tem sido publicado um grande número de estudos sobre as experiências com o ensino remoto (emergencial), referentes a diferentes contextos educativos (escolar, universitário e formação contínua). Estes contributos são, frequentemente, estudos de caso, muitas vezes baseados em questionários, focando aspetos como competências digitais (Bezerra, 2021; Petzold-Rudolph *et al.*, 2023), uso de ferramentas digitais no ensino remoto (Strömsdörfer, 2021; Bonifácio & Siqueira, 2023), desafios técnicos (Alolaywi, 2021; Leite *et al.* 2021), implicações nas relações sociais em contextos virtuais (Faria & Almeida, 2023), bem-estar mental e psicológico (Bezerra, 2021; Flores *et al.* 2021) bem como a eficiência do ensino remoto (Mishra, Gupta & Shree, 2020; Cutri, Mena & Whiting, 2020; Bao, 2020), concentrando-se, ora na perspetiva do docente, ora na perspetiva do discente.

No seu estudo a grande escala, Traxler (2023) analisou 278 publicações europeias sobre o impacto que a súbita passagem ao ensino remoto teve nas atividades letivas das universidades, elencando algumas recomendações, como por exemplo um melhor provisionamento técnico das salas de aula bem como uma maior flexibilização de estruturas curriculares através de mais ofertas de ensino a distância. Diante dos novos desafios da digitalização, sugere ainda a implementação de medidas para aprimorar as competências dos docentes e, conseqüentemente, a qualidade do ensino superior. Isso exige uma qualificação específica dos professores, especialmente na área da didática voltada para o ensino remoto. Após analisar os estudos publicados sobre o ensino remoto, conclui que a necessidade de uma metodologia pedagógico-didática adaptada a esse formato foi o aspeto menos abordado. Alerta que os métodos utilizados no ensino presencial não são diretamente transferíveis para o ensino digital, o qual exige a aplicação de novas abordagens para a transmissão de conhecimentos e o

desenvolvimento de competências específicas para o ambiente de aprendizagem *online*.

Theelen & Breukelen (2022) apresentam a análise de 42 artigos, escolhidos segundo critérios definidos, a saber: redação em inglês, *peer-review*, método qualitativo, quantitativo ou misto, contendo palavras-chave relacionadas com didática e pedagogia do *e-learning* no ensino superior (p. 1289). Os autores constataam que, apesar da grande quantidade de artigos publicados, pouco se sabe sobre a didática específica e os princípios pedagógicos do design adaptado ao *e-learning*, embora este se tenha tornado, sensivelmente desde 2010, cada vez mais presente na prática letiva. Tal como Traxler (2023), lamentam que os estudos mais recentes revelem que a tecnologia em educação seja usada apenas para substituir as aulas presenciais tradicionais (*face-to-face*) (p. 1287), quando a tecnologia oferece um grande potencial para enriquecer o ensino, criando tarefas de ensino-aprendizagem antes inconcebíveis. Sublinham ainda a falta de pesquisa bem fundamentada para entender melhor os princípios de conceção didática e pedagógica do *e-learning*. Definem a terminologia-chave utilizada no artigo (didática, pedagogia, *e-learning*, *blended learning*).⁸ Os autores propuseram a si mesmos o objetivo ambicioso de fornecer instrumentos para preencher esta lacuna, dando alguns conselhos básicos sobre o *e-learning*. Assim, a análise dos artigos selecionados leva à compilação de quatro princípios, devidamente definidos, que devem ser contemplados no *e-learning*: “(1) content scaffolding, (2) process scaffolding, (3) peer-to-peer learning, and (4) formative strategies” (p. 1286). Apesar de terem conseguido identificar algumas semelhanças nos artigos analisados, tais como a necessidade de um equilíbrio entre estratégias centradas no docente e nos estudantes, assim como o objetivo de aumentar a autonomia destes, chegam à conclusão de que não existe nenhum “ready-to-use recipe” (p. 1297), além de que

⁸ “*E-learning* can take place on a continuum from real-time interaction (synchronous) to fully asynchronous interaction. For example, when using a chat or videoconference, online interactions can take place real time. On the contrary, when using email or discussion boards, interactions are asynchronous.” (p. 1288). Referem a definição de Torrisi-Steele (2011, p. 366): “Blended learning refers to enriched, student-centred learning experiences made possible by the harmonious integration of various strategies, achieved by combining face-to-face interaction with ICT.” (p. 1287) (ICT = Information and communications technology).

os princípios fundamentais caracterizando o *e-learning* não são exclusivos desse tipo de ensino.

Kergel & Heidkamp-Kergel (2020) apresentam um modelo para a conceção de novas formas de ensino e aprendizagem digitais, incluindo várias etapas, a saber: análise, planeamento, desenvolvimento, implementação e avaliação (p. 16). Contudo, parece-nos evidente que o modelo proposto precisaria de ser adaptado a diferentes contextos educacionais.

Quanto à didática e à pedagogia específicas das aulas *online* de línguas estrangeiras, que se caracterizam, intrinsecamente, pela já referida grande componente de interação oral, o panorama é ainda mais precário, visto que faltam contributos tratando a didática específica necessária. Existem sim, inúmeros estudos de caso sobre o ensino das línguas em tempos de confinamento. Destes contributos que tematizam aspetos como fatores psicológicos, sociais e emocionais (distância social, prazer e satisfação na aprendizagem), eficiência das aulas online, interação docente-discente e entre discentes, só podemos referir alguns, a título de exemplo: Klimova, 2021: ensino de línguas estrangeiras na Universidade de Hradec Kralove, República Checa; Toomnan, 2024: ensino do inglês, numa universidade pública tailandesa; Colombo & Boscia, 2022: ensino do francês e italiano, numa universidade da Malásia; Cocoradă, & Maican, 2021: ensino de línguas estrangeiras no ensino superior.

Em suma, a revisão da literatura mostra que, embora haja pontos de contacto relativos ao modo como o ensino remoto é concebido e percecionado, faltam modelos didáticos concretos direcionados para o ensino das línguas estrangeiras com as suas especificidades. Além disso, as diferenças contextuais significativas exigem que as propostas para abordagens pedagógico-didáticas também sejam específicas.

3. Metodologia

O procedimento metodológico usado neste projeto assentou no modelo de investigação qualitativa (cf. Kuckartz, 2008, 2014, 2018; Flick, 2011; Mayring, 2014). Para a seleção dos docentes inquiridos seguiu-se o princípio de “purposive sampling”

de Silverman (2010, p. 141), visando constituir um grupo de participantes das oito universidades que, pela sua vasta experiência letiva, asseguraram um alto teor informativo, a fim de evitar generalizações empíricas (Patton, 2015, p. 264). Realizaram-se dez entrevistas com os docentes. Partindo das questões de investigação, elaborou-se um guião de entrevista semiestruturada (Riemer, 2016; Flick, 2011), permitindo um determinado grau de flexibilidade para explorar as experiências individuais dos entrevistados (Riemer, 2016, p. 163). Após uma fase de teste do guião, contendo as perguntas orientadoras, as entrevistas, de duração média de 60 a 90 minutos, foram realizadas e gravadas por Zoom e posteriormente transcritas, seguindo os princípios de transcrição de Dresing & Pehl (2018). Nas subseqüentes publicações de resultados parciais do estudo, os dados referentes a universidades e docentes foram anonimizados, eliminando características identificadoras dos dados de pesquisa e atribuindo códigos, tanto aos docentes (D1, D2, D3, etc.) como às instituições de ensino (U1, U2, U3, etc.) (Meyermann & Porzelt, 2014, p. 4). Para a análise dos dados empíricos assim obtidos aplicou-se o método de análise qualitativa de Kuckartz (2018), recorrendo ao suporte do software Maxqda 2021 (Verbi Software, 2019)⁹. Esta análise qualitativa assistida por computador parte da constituição de categorias temáticas e da definição indutiva de subcategorias para posterior descrição, interpretação e consolidação de resultados (Kuckartz, 2018, p. 100).¹⁰

O segundo grupo de inquiridos, os estudantes de alemão frequentando cursos de licenciatura das mesmas universidades ($n = 156$), preencheu um questionário realizado em formato online, baseado em perguntas com respostas pré-definidas, com inclusão de algumas perguntas abertas. Este questionário, antes de ser aplicado, percorreu duas fases de *pretest*. O inquérito foi anónimo, permitindo, contudo, através da inclusão de informações referentes a idade, sexo, universidade, curso de licenciatura (e ano de frequência do mesmo), estabelecer correlações relevantes entre as universidades participantes, no que diz respeito à conceção do ensino *online* por parte das universidades e dos docentes, e à receção do ensino remoto por parte dos estudantes.

⁹ <https://www.maxqda.com>

¹⁰ Os segmentos codificados com Maxqda foram numerados, permitindo a atribuição e identificação de cada segmento. D4, 45 significa, por exemplo, docente inquirido 4, segmento 45.

Assim, a realização do projeto decorreu, sucessivamente, nas seguintes etapas. Na primavera de 2021, foram realizadas as entrevistas com os docentes, posteriormente transcritas. Partindo da transcrição, definimos categorias e subcategorias e passou-se à codificação dos segmentos. No final do ano letivo de 2021/22, realizou-se o inquérito aos estudantes. A última fase do projeto consistiu na análise dos dados e na apresentação ou publicação de resultados parciais.

FIGURA 1 - Estrutura do projeto

Estrutura do projeto – grelha orientadora		
	Docentes	Estudantes
Contexto	Alemão como língua estrangeira nas universidades portuguesas	
Abordagem da investigação	A. Conceção do ensino digital B. Implementação de novos formatos de ensino C. Efeitos de fatores exteriores (económicos, sociais, emocionais) para o funcionamento do ensino	A. Receção do ensino digital B. Aceitação de novos formatos de ensino/aprendizagem C. Efeitos de fatores exteriores (económicos, sociais, emocionais) para a aprendizagem
Método: Análise qualitativa (descritiva) (Kuckartz 2018) → codificação e análise de dados com MAXQDA	10 entrevistas por tópicos (perguntas abertas) com docentes de alemão como língua estrangeira em 8 universidades portuguesas públicas 1ª fase: Realização das entrevistas 2ª fase: Transcrição e codificação 3ª fase: Análise dos dados	Questionário online estandardizado, com perguntas semi-abertas (em 8 universidades portuguesas públicas) 1ª fase: Pretest 2ª fase: Realização do questionário nas universidades (em curso) 3ª fase: Análise dos resultados
Objetivos	<ul style="list-style-type: none"> • Identificação de estratégias aplicadas: boas práticas • Elaboração de modelos de ação para a conceção de formatos online de ensino/aprendizagem do alemão • Desenvolvimento de modelos sustentáveis e recomendações exequíveis na prática letiva para a era pós-pandémica → integrando as boas práticas de ensino online, em contexto de ensino presencial (independentemente de medidas de confinamento/Covid-19) 	

(Fonte: elaborado pelas autoras)

4. Resultados obtidos e análise de dados

4.1. Março de 2020: passagem repentina ao ensino *online*

A passagem repentina para o ensino a distância foi marcada pela necessidade de mudanças rápidas, muitas vezes improvisadas, e constituiu, para a maioria dos inquiridos, um enorme esforço suplementar. Em quase todas as entrevistas surgiu repetidamente o termo “Überforderung” (termo alemão que define uma situação de sobrecarga em que a pessoa não se sente capaz de resolver

ou ultrapassar adequadamente um problema, uma situação ou um desafio, com as competências e recursos de que dispõe, atingindo assim os seus limites):¹¹

Essa transição do ensino presencial para o ensino *online* [...] foi muito surpreendente e todos nós nos sentimos sobrecarregados [...] Tive a sensação de que, no início, quisemos fazer tudo o que era possível. Por isso, também usamos todas as aplicações a que tínhamos acesso. Usamos o Panopto para gravar vídeos [...] Usamos o Moodle para enviar tarefas e como plataforma onde os textos eram entregues e os vídeos do Panopto eram carregados. [...] Tive a sensação de que tentamos utilizar tudo o que nos foi disponibilizado pela universidade. (D2, 15)

4.2. Oferta de formações dedicadas às plataformas de ensino *online*

Em quase todas as universidades houve uma grande oferta de formações e *workshops* para aprender as funcionalidades e potencialidades do Zoom ou Moodle, mas também para funções mais complexas como a elaboração de testes. Em algumas universidades houve a possibilidade de participar em formações para trabalhar com outras aplicações, por exemplo Kahoot ou ainda Panopto, e para elaborar vídeos didáticos. É de notar ainda que, apesar de haver estas ofertas por parte das universidades, levou muito tempo e dedicação até os docentes estarem suficientemente familiarizados com as novas aplicações e serem capazes de planear e realizar uma unidade letiva com as mesmas.

Contudo, houve exceções. Um docente queixou-se do facto de ter havido apenas uma formação, num só dia, logo no início de março de 2020, mas, como nem todos os docentes tiveram a disponibilidade de participar, foi necessário adquirir as competências mais tarde, sozinho ou com a ajuda de tutoriais rudimentares:

Recebemos um e-mail sobre o funcionamento do Zoom logo no início, em março. [...] houve cinco *screenshots* sobre o funcionamento do Zoom, mas as imagens dos botões a premir estavam esbatidas. Por isso, era basicamente "salve-se quem puder." (D3, 24)

¹¹ Todas as citações das entrevistas com os docentes foram traduzidas para português e, para melhor compreensão, ligeiramente corrigidas, eliminando repetições, frases inacabadas, erros gramaticais, etc.

Outro inquirido lamentou não ter havido formações em tempo real, porque tudo era disponibilizado apenas em formato de tutorial para o estudo autónomo:

Houve apoio pela universidade, mas era tudo por tutoriais, ou seja, por links que podíamos explorar por nós próprios. Infelizmente, não houve sessões ao vivo. Era tudo por correio eletrónico e com instruções, não houve nenhuma formação. (D9, 19)

Repetidamente, os docentes inquiridos manifestaram o seu desagrado pelo facto de a sua faculdade ter disponibilizado formações gerais dedicadas às ferramentas digitais Zoom e Moodle, sem que tenha sido oferecida qualquer formação didática para a passagem ao ensino *online*.

Tanto quanto sei, não houve [cursos de formação] didáticos. Particpei em duas ações de formação técnica sobre o Moodle. Essa foi via Zoom, e houve várias formações sobre o Moodle. (D6, 41-42)
E também não houve formação didática. (D7, 43)

Foi criticado também o facto de as formações se dirigirem indiscriminadamente aos docentes de todas as faculdades, não contemplando as especificidades de cada área. O que pode ser um instrumento ou uma estratégia útil para as ciências não é necessariamente aplicável nas aulas interativas de línguas estrangeiras:

São formações que estão abertas a todos os docentes da universidade, incluindo colegas das ciências, de medicina, etc. E quando vejo quem participa nestas formações, são maioritariamente docentes das ciências humanas. [...] E como as formações são feitas, são muito centradas nas ciências e nos testes em massa, para centenas de estudantes que fazem todos o mesmo teste. E nós temos uma forma de avaliação muito individualizada. (D7, 44-45)

Por isso, foi interessante notar que os docentes das universidades 8 e 9 referiram que as formações oferecidas por editoras especializadas na área do ensino de alemão como língua estrangeira e pelo Goethe-Institut (Instituto Alemão) foram mais úteis do que as formações oferecidas pela universidade:

As editoras de manuais de alemão como língua estrangeira, Klett e Hueber, por exemplo. O Instituto Goethe [Instituto Alemão] também tinha uma série de ofertas formativas. Participava nelas de vez em quando. Do ponto de vista científico, eram definitivamente mais enriquecedoras do que as ofertas da universidade. (D8, 42-43)

Mas como também trabalho para o Instituto Goethe e lá tivemos formações, adquiri [as competências necessárias] noutra sítio. (D9, 20-21)

Os problemas técnicos, como a instabilidade da Internet e a necessidade de utilizar equipamentos próprios, foram desafios recorrentes. “No início da pandemia houve muitos problemas com o Zoom, nessa altura sentia falta de apoio técnico.” (D3, 30). Além das limitações estritamente técnicas, também constatamos queixas por perder tempo e dinâmica nas aulas *online* (D1, 44). Os problemas técnicos relatados pelos docentes inquiridos já foram descritos numa outra publicação nossa.¹²

4.3. Reação dos docentes

Diante da emergência educacional provocada pela pandemia, a flexibilidade e o poder de reação rápida dos docentes foram cruciais para a adaptação ao novo contexto de ensino. Na universidade 4, por exemplo, os docentes de línguas estrangeiras desenvolveram iniciativas próprias para reagir à situação excepcional, mobilizando sinergias das várias áreas científicas. Colegas que já tinham experiência com o ensino a distância por terem lecionado módulos em formato de *blended learning* partilharam esta experiência com colegas de outras áreas:

Na nossa faculdade, há mais de dez anos, temos um programa de estudos em Filosofia, que foi concebido e realizado em *blended learning*. Isso significa que já havia alguns docentes que estavam familiarizados, por exemplo, com o Zoom. Essa experiência foi muito importante, porque, como mencionei, na tarde do nosso último dia presencial na universidade, em março de 2020, tivemos uma espécie de reunião de crise na reitoria e entramos em contacto com esses três colegas que tinham larga experiência com ferramentas digitais no ensino. Eles prepararam

¹²Comunicação (publicação no prelo) por Kind, A, Herget, K, Scholz, I.: “Projeto de Investigação inter-universitário “Digi-DaF” (Parte 1): Desafios e Oportunidades de novos formatos digitais no Ensino-Aprendizagem do Alemão na Perspetiva dos Docentes”, *III International Conference Aprolínguas 2021: “Foreign Language Teaching and Learning in the Classroom and online: Tradition, Perspectives and Challenges”*, 2-3.12.2021.

rapidamente um breve tutorial e até mesmo um módulo para o ensino *online*. (D4, 43-44)

Na universidade 8 também houve partilha de experiências entre docentes, mas não se revelou tão organizada e proveitosa como no exemplo da universidade 4, acima referido:

A aplicação Padlet foi explicada com mais detalhes por uma colega, e também houve uma introdução ao trabalho com aplicações de colegas de outros departamentos que não são muito interessantes para nós. É sempre preciso fazer uma seleção. Também foi tema a elaboração de testes *online*, o que ocupava muito tempo. Mas para além do Padlet, não me lembro se foram apresentadas outras ferramentas que sejam relevantes ou interessantes para nós. (D8, 40-41)

O docente 4 também enfatizou que os docentes podem aprender muito pelas práticas didáticas na área das outras línguas estrangeiras:

Como temos unidades curriculares de inglês, português, espanhol, francês, alemão, italiano, português como língua estrangeira, bem como línguas exóticas como japonês e chinês, podemos realmente aprender muito uns com os outros. [...] No primeiro semestre deste ano, eu e alguns colegas organizamos uma formação interna sobre *best practices* nas diversas áreas de línguas. [...] Foi muito interessante ouvir o que os outros colegas estavam a fazer, e partilhar ideias de forma informal. [...] E agora estamos em vias de realizar um segundo encontro interno, para docentes de línguas, porque alguns dos nossos colegas até têm o doutoramento na área de didática de línguas estrangeiras, o que significa que estão muito mais informados sobre a matéria. Não possuem apenas experiência, mas também conhecem a literatura da especialidade. (D4, 48)

Este docente argumentou ainda que a disposição dos colegas para colaborar intensamente e partilhar experiências foi impulsionada pela falta de contacto social resultante do confinamento.

Este aspeto social faz falta ao fim de algum tempo. Acho que isso também contribuiu para o facto de todos os colegas estarem tão dispostos a trabalhar em conjunto. (D4, 48)

De uma maneira geral, o ensino a distância afetou negativamente a vida social e o bem-estar emocional de docentes e estudantes, reduzindo o contacto dos estudantes entre si e com os seus docentes e impossibilitando as atividades

culturais presenciais nas faculdades. A restrita interação nas aulas *online*, às vezes agravada por câmaras desligadas, contribuiu para um ambiente distante e, em alguns casos, as restrições do confinamento resultaram em solidão e estados depressivos.

4.4. Ensino de línguas em aulas *online*: especificidades, desafios e *workload*

Os aspetos didáticos do ensino a distância revelaram-se desafiadores, especialmente no desenvolvimento da expressão oral, devido à falta de interação visual e auditiva de qualidade (D3, 30). A falta de contacto direto e a impossibilidade de observar a linguagem corporal dos estudantes dificultaram a avaliação precisa das suas competências comunicativas. As quatro competências linguísticas (ouvir, falar, ler e escrever) foram afetadas de maneiras diferentes, mas a competência de falar provou ser a mais problemática.¹³

Especificamente no ensino de línguas, essas dificuldades são ainda mais pronunciadas em comparação com disciplinas teóricas. A componente comunicativa e interativa é essencial para a aprendizagem eficaz de uma nova língua. Num ambiente de sala de aula tradicional, os estudantes podem praticar a fala através de atividades em grupo e interações diretas, o que não é tão facilmente replicável *online*. Os docentes tiveram de encontrar maneiras criativas para incentivar a interação, como o uso de salas simultâneas (*breakout rooms*) para atividades orais em pequenos grupos. Porém, alguns dos docentes inquiridos constataram que não podiam estar em todos os grupos ao mesmo tempo e, assim, não tinham controlo sobre o que acontecia nas salas simultâneas (se os estudantes colaboravam de forma equitativa ou se falavam, de facto, na língua-alvo).

Em suma, enquanto as salas simultâneas proporcionaram uma oportunidade para promover a interação e a prática oral em pequenos grupos, a falta de controle direto por parte dos docentes evidenciou as limitações e a necessidade contínua de inovação e adaptação nas metodologias de ensino a distância. Ou seja: não

¹³Auf der Maur, S., Hergel, K. & Kind, A. (2024), Die Fertigkeit Sprechen in der universitären Online-DaF-Lehre zu Zeiten der Covid-19-Pandemie: Herausforderungen und Desiderata. *Atas do I Simpósio Internacional de Língua Alemã nos Países de Língua Oficial Portuguesa*, São Tomé e Príncipe, 14-16 junho de 2023, no prelo.

basta dominar as tecnologias oferecidas por plataformas como o Zoom, se não se dispõe de estratégias como fazer um trabalho eficiente com as mesmas.

Outro docente relatou que tudo que, numa aula presencial convencional, se anota no quadro de forma espontânea ao longo da aula, no ensino *online* tinha de ser preparado em diapositivos PowerPoint, o que levou muito tempo. Estas apresentações em PowerPoint tiveram de ser feitas de tal modo que pudessem ser publicadas no Moodle ou noutras plataformas para serem consultadas em qualquer momento, sobretudo por estudantes que não tinham tido a possibilidade de assistir às aulas, já que a obrigatoriedade de presença foi suspensa durante as fases de confinamento. Outro docente destacou o acréscimo de trabalho resultante da apropriação das novas ferramentas digitais.

A disponibilização de materiais didáticos no Moodle foi essencial para o sucesso do ensino a distância. No entanto, a integração de aplicações em aulas *online* variou significativamente, dependendo da afinidade e experiência prévia dos docentes com esses recursos e, em alguns casos, da sua faixa etária. Os docentes mais jovens tenderam a utilizar mais aplicações, mas o benefício didático dessas ferramentas é questionável e precisa ser cuidadosamente avaliado. As limitações das aplicações foram destacadas em vários relatos de docentes durante as entrevistas. Estas aplicações tendem a fragmentar as competências linguísticas, focando-se principalmente em microestruturas, como género, formação do participio passado e léxico. Contudo, falham na abordagem de funções comunicativas mais amplas, que envolvem a construção de frases inteiras e a realização de enunciados completos para cumprir funções comunicativas específicas. Também o fator lúdico foi mencionado, pois rapidamente pode transformar-se numa distração em vez de um momento de aprendizagem.

Um aspeto que é destacado em quase todas as entrevistas é o tempo dispensado para a correção de composições e outros trabalhos escritos. A correção em formato digital, seja com a função de comentários em Word, seja por outros tipos de destaques de correções e comentários pelo docente, exige muito mais tempo do que a correção à mão de textos em papel, manuscritos ou impressos. O trabalho do estudante de alemão tem de ser consultado, seja qual for o canal pelo qual é disponibilizado, tem de ser guardado numa pasta previamente preparada,

com uma designação feita pelo docente segundo o método que escolheu, é corrigido, erro por erro, o que, consoante o método da correção, exige vários cliques por cada erro, é guardado na versão corrigida e comentada, e finalmente é disponibilizado ao estudante (D4, 91-92).

Em algumas faculdades os docentes corrigem segundo um modelo didático que exige uma primeira e uma segunda versão, ambas sendo completadas por siglas de erro ou corrigidas, respetivamente. Assim, o docente 7 indica a correção de textos em formato digital como a tarefa mais dispendiosa em termos de tempo.

A maior carga de trabalho era a correção, porque me enviavam os trabalhos, e acho que o pior era corrigir os textos individuais de uma aula para a outra, porque não conseguia dar conta. Chegava à aula seguinte sem ter ainda corrigido os textos da aula anterior, mas tinha de continuar e, por isso, foi-se acumulando o trabalho, e, a certa altura, ultrapassou as minhas capacidades, digamos assim. (D10, 75)

Estou-me a referir aos trabalhos de casa, se todos os estudantes fizerem duas ou três versões dos textos, a correção demora muito tempo. Cheguei a olhar para o relógio. Se eu corrigir à mão, acho que consigo fazer o mesmo num terço do tempo. São precisos quinze cliques para abrir um documento, guardá-lo e assim por diante. É inacreditável, consome imenso tempo. (D7, 71-72)

Com a transição para o ensino *online*, também surgiu a necessidade de adaptar métodos tradicionais de avaliação, incluindo testes escritos e provas orais, para o formato digital. Embora a implementação de provas *online* tenha permitido a continuidade das avaliações durante a pandemia, vários desafios foram identificados, revelando que essas provas não são diretamente comparáveis com os testes realizados de forma presencial.

Os docentes constataram que o tempo necessário para a aplicação e a conclusão de provas em formato digital é significativamente superior ao dos testes presenciais. Isso deve-se a diversos fatores, incluindo dificuldades técnicas, a necessidade de adaptação dos estudantes ao ambiente virtual e a definição de condições para a realização das provas, tentando evitar fraudes e garantir justiça e transparência.

A correção das provas realizadas *online* também apresentava grandes desafios. Embora existam plataformas (por exemplo, o Moodle) que oferecem funções de teste com correção automática, essas ferramentas têm limitações

significativas. Estas modalidades de correção são adequadas para testar competências fragmentadas, em situações em que as respostas podem ser claramente categorizadas como certas ou erradas. Estas funções de teste podem ser muito úteis para outras áreas científicas. Contudo, no caso das línguas estrangeiras, esse tipo de teste não é adequado, pois as respostas pedidas devem ser mais complexas, e não só no caso da produção textual (redações e ensaios). A avaliação de competências linguísticas exige uma análise mais detalhada e contextual, que vai além da simples verificação de respostas certas ou erradas. A capacidade de um estudante de formular argumentos coerentes, estruturar o seu discurso, utilizar vocabulário apropriado e demonstrar fluência na escrita não pode ser adequadamente medida por correção automática.

Resumindo, pode constatar-se que os docentes adquiriram com relativa rapidez as competências básicas necessárias para lecionar uma aula em formato digital (Zoom, Teams, Moodle, etc.), mas sentiram a falta de formações específicas destinadas a demonstrar como praticar e desenvolver, de forma eficiente e produtiva, as quatro competências (falar, ouvir, ler e escrever), no novo formato digital:

Gostaria muito de fazer um curso básico de ensino a distância, porque eu própria tenho muitas dúvidas, e gostaria de saber [as respostas], porque tenho a certeza de que há pessoas que estão a fazer pesquisa nesta área. (D3, 58)

Senti um pouco a falta da parte didática. Tive a sensação de que estávamos completamente entregues a nós próprios. Qual será o melhor método? Qual a melhor forma de chegar aos estudantes? Penso que isso se deveu principalmente ao facto de ninguém se ter debruçado sobre essas questões antes. Ensinar *online* até à data não tinha sido assunto, por isso fomos todos atirados aos leões, por assim dizer, mas teria sido bom se tivéssemos também tido apoio didático: como trabalhar ou ensinar melhor com ferramentas *online*? Quanto à formação técnica [Zoom, Moodle], houve oferta suficiente por parte da universidade. (D2, 33)

Em suma, durante o período de transição abrupta para o ensino *online* em resposta à pandemia, os docentes enfrentaram uma sobrecarga significativa de trabalho adicional. Este incluiu a familiarização com novas aplicações, a adaptação de materiais didáticos para o ensino a distância e a correção demorada de trabalhos em formato digital.

Os docentes inquiridos demonstraram grande empenho e capacidade de adaptação perante esses desafios. Utilizaram uma variedade de ferramentas digitais, como o Moodle, o Zoom, o Panopto, o Kahoot e o Mentimeter, e disponibilizaram materiais suplementares *online*. Além disso, tiveram de despender muito tempo em atendimentos individuais e na resposta a uma quantidade inusual de e-mails, fazendo um esforço adicional para dar apoio aos estudantes e manter o contacto com os mesmos. Apesar do esforço contínuo e da colaboração mútua, as limitações técnicas e o *workload* adicional frequentemente comprometeram a eficiência didático-pedagógica. Ficou evidente, na análise das entrevistas, que os docentes estavam a enfrentar uma situação de crise e que, quando utilizaram os recursos disponíveis, o faziam de forma espontânea e não de forma previamente planeada e pedagogicamente refletida. O grande esforço por parte dos docentes que salvaguardou o ensino remoto emergencial foi também destacado por outros autores, como, por exemplo, Mavridi (2022), que conduziu um estudo sobre as experiências feitas por docentes de inglês ($n=1102$) durante o Covid-19. Este estudo mostra que, no início da pandemia, a formação dos docentes não era suficiente para enfrentar os desafios que o ensino à distância trouxe:

The findings of this research show that language educators worldwide exhibited remarkable resilience and innovation during the first phases of their shift to Emergency Remote Teaching (ERT). Through trial and error, they showed determination as they stepped out of their comfort zones and adapted to new teaching modalities. (Mavridi, 2022, p. 43)

Pudemos constatar um notável trabalho colaborativo na Universidade 4 como *best practice*, onde os docentes desenvolveram sinergias, aproveitando as suas experiências e organizando, por iniciativa própria, uma série de *workshops* de introdução ao trabalho com ferramentas digitais e de partilha de boas práticas no ensino *online* das línguas estrangeiras e de experiências com o ensino a distância ou *blended learning*.

Porém, o facto de termos constatado, após análise dos dados recolhidos, mais manifestações de problemas e dificuldades por resolver do que *best practices* a compilar, motivou a equipa de investigação a reformular os objetivos do projeto, insistindo mais na definição de *desiderata* do que na proposta de boas práticas.

Assim, o título do projeto de investigação passou a ser: "Digi-DaF: Desafios do ensino universitário digital de línguas, pelo exemplo do ensino do alemão como língua estrangeira, durante a pandemia Covid-19 e posteriormente: boas práticas e *desiderata*".

5. Um olhar para o futuro: *desiderata*

Sem querer agourar: a próxima pandemia virá mais rapidamente do que queríamos, só não sabemos quando. A consciência da possibilidade de futuras pandemias reflete-se em iniciativas tomadas por várias organizações, das quais vamos referir duas, a título de exemplo. A EUGLOHRIA (European University Alliance for Global Health) tem vindo a organizar eventos académicos que visam partilhar ideias e resultados de pesquisa científica referentes a desafios decorrentes do Covid 19 e futuras pandemias.¹⁴ O segundo exemplo a destacar é a iniciativa dos estados-membros da Organização Mundial de Saúde (OMS) que estão, neste momento, a negociar as linhas condutoras de um acordo global de pandemia. Como consta na publicação da OMS que apresenta estas negociações, para muitos cientistas é um facto consumado que iremos enfrentar uma próxima pandemia, seja ela causada por um vírus da gripe, pelo SARS-Cov-3, pelo vírus da varíola ou por bactérias.¹⁵

Nesse cenário, parece-nos que não só o sistema de saúde devia estar mais bem preparado, como também o sistema educativo. Ou seja, os governos e as universidades deviam, neste momento, estar a debruçar-se intensamente sobre o desenvolvimento de mecanismos alternativos ao ensino presencial, em escolas e universidades, para estarem mais bem preparados para a próxima situação de crise.

No caso específico do ensino universitário das línguas estrangeiras, devemos compilar, valorizar e divulgar as boas práticas, mas também aprender com os erros cometidos, a fim de podermos enfrentar uma nova crise com estratégias acertadas. Partindo da análise dos dados obtidos no nosso estudo, apresentamos as seguintes recomendações:

¹⁴ Universidade do Porto: "Workshop on Developing Guidelines for Future Pandemics" (25/10/2023). <https://www.euglohria.eu/r-and-i-in-action/euglohrias-approach-and-efforts>.

¹⁵ <https://www.who.int/news/item/28-05-2024-who-member-states-agree-way-forward-to-conclude-pandemic-agreement>.

- Elaboração de um plano para a formação contínua dos docentes de línguas estrangeiras nas universidades portuguesas para a aquisição e aprofundamento de competências digitais transversais;
- Formações dirigidas aos estudantes das universidades portuguesas, para a aquisição e o aprofundamento de competências digitais transversais;
- Formação específica em didática de ensino *online* para docentes de línguas estrangeiras;
- Disponibilização de equipamento para todos (como por exemplo computadores portáteis), Internet estável;
- Disponibilização de salas adequadamente equipadas para modelos híbridos em que os estudantes também possam falar sem incomodar os outros (no caso de os modelos híbridos serem parcialmente presenciais e parcialmente *online*, no mesmo dia);¹⁶
- Regulamento a nível nacional para o ensino a distância (uniformização das medidas de proteção de dados: câmara ligada em sessões *online*?);
- Disponibilização de recursos humanos nas universidades portuguesas para se dedicarem à área da didática do ensino *online* e para desenvolverem estratégias e abordagens metódico-didáticas que, posteriormente, possam ser partilhadas com os colegas;
- Incentivos para o trabalho colaborativo entre docentes das várias línguas estrangeiras e/ou outras áreas científicas.

Em conclusão, a implementação de planos contínuos de formação em competências digitais, juntamente com a disponibilização de recursos adequados e incentivos para os docentes, é essencial para modernizar e melhorar o ensino de línguas estrangeiras nas universidades portuguesas. Além disso, a criação de

¹⁶ Os vários modelos híbridos praticados nas universidades portuguesas durante o confinamento em 2020 já foram apresentados na comunicação (publicação no prelo):

Auf der Maur, S, Riedel, A., Ascher, C.: „Interuniversitäres Forschungsprojekt "DIGI-DAF" (Teil 2): Modelle und Methoden des Fremdsprachenunterrichts in Zeiten von Covid-19 - Herausforderungen und Chancen". III *International Conference Aprolínguas 2021: "Foreign Language Teaching and Learning in the Classroom and online: Tradition, Perspectives and Challenges"*, 2-3.12.2021.

regulamentações claras para o ensino a distância e o fomento do trabalho colaborativo entre diferentes áreas científicas promoverão uma abordagem educacional mais integrada e eficaz, e assim preparada para o futuro. Compete não só às universidades criar as condições imprescindíveis para uma eficiente e proveitosa realização do ensino a distância, mas também aos docentes de línguas estrangeiras organizarem-se e tomarem a iniciativa para a partilha dinâmica de experiências e a elaboração de estratégias adequadas.

Considerações finais

Neste artigo propusemo-nos explorar a formação oferecida pelas universidades para assegurar a transição repentina do ensino presencial para o remoto. Além disso, analisamos os desafios enfrentados pelos docentes universitários de alemão na busca de soluções adequadas e no esforço para desenvolver estratégias e procedimentos metodológicos adaptados ao ensino digital. A análise desses aspetos permitiu avaliar o grau de preparação pedagógico-didática dos docentes frente aos desafios do ensino remoto.

É fundamental lembrar que, com o surto da pandemia, o ensino a distância teve de ser implementado de forma abrupta e que, em geral, nem as universidades, nem os docentes, nem os estudantes tinham experiência suficiente com essa modalidade de ensino. A análise dos dados revelou que os docentes inquiridos tentaram dar o seu melhor para encontrar um caminho para adaptar as suas aulas ao formato *online*. O estudo de Klimova (2021), baseado num inquérito aos estudantes, corrobora esta constatação: “[...] they [students] admit that teachers really try hard to make the online foreign language teaching stimulating” (p. 1791). Os docentes inquiridos no nosso estudo viram-se confrontados com a falta de bases didáticas específicas e de formações adequadas para as suas necessidades. Assim, as abordagens adotadas não foram tão bem planeadas e concebidas, nem com tanta antecedência, como no ensino presencial.

Podemos concluir que, basicamente, não cabe ao docente de línguas estrangeiras elaborar individualmente um formato adequado para o ensino *online*,

mas sim a uma equipa de especialistas que se dedique ao desenvolvimento da didática para esse tipo de ensino. Foi isso que faltou - uma base metódico-didática do ensino a distância adaptada às especificidades do ensino das línguas estrangeiras, no nosso caso, do alemão. Assim, as experiências feitas por docentes durante a pandemia, no fundo, representam apenas o início de processos de desenvolvimento institucional a longo prazo.

A análise dos dados obtidos e da literatura da especialidade mostrou que faltam contributos e pesquisa específica para a didática e para a pedagogia do ensino remoto em geral, e das línguas estrangeiras em particular. Theelen & Breukelen (2022) confirmam essa observação: "Although much has been written about *e-learning*, little is known about crucial didactic and pedagogical design principles for *e-learning*" (p. 1286). Charlier *et al.* (2015) também destacam essa lacuna, constatando que os actantes envolvidos (instituições de ensino, docentes e estudantes) dispõem de pouca pesquisa científica, estando limitados, essencialmente, a relatórios de experiências, casos de estudo e questionários de satisfação.

Além da realização de mais pesquisa de base, o cenário ideal, para o ensino superior estar preparado para situações que obriguem à passagem forçada ao ensino remoto, implicaria a formação de uma equipa de docentes de línguas estrangeiras, especialmente destinada à elaboração de estratégias pedagógico-didáticas, apoiada pela universidade, não só logisticamente, mas também pela disponibilização de tempo para a realização dessa tarefa, criando assim uma sinergia produtiva das medidas implementadas pelas universidades e da colaboração dos seus docentes.

Referências

- Alolaywi, Y. (2021). Teaching online during the COVID-19 pandemic: Teachers' perspectives. *Journal of Language and Linguistic Studies*, 17(4), 2022-2045.
- Bao, W. (2020). COVID-19 and online teaching in higher education: A case study of Peking University. *Human Behavior and Emerging Technologies*, 2(2), 113-115. <https://doi.org/10.1080/02619768.2020.1815702>.
- Bezerra, F. A. S. (2021). Experiências de ensino-aprendizagem remoto de inglês na licenciatura em Letras/Inglês durante a pandemia de COVID-19:

- Multiletramentos digitais e interseccionalidade. *Revista da Universidade Federal da Paraíba*, 10(2), 41-66.
- Bonifácio, L., & Siqueira, G. (2023). Metodologias e métodos de ensino de línguas durante a pandemia COVID 19: Uma revisão de literatura entre 2020 e 2022 em periódicos nacionais. *Revista Desempenho*, 1(33), 86-106. <https://periodicos.unb.br/index.php/rd/article/view/44523>.
- Charlier, B., Cosnefroy, L., Jézégou, A., & Lameul, G. (2015). Understanding Quality of Learning in Digital Learning Environments: State of the Art and Research Needed. In A. Curaj, L. Matei, R. Pricopie, J. Salmi, & P. Scott (Eds.), *The European Higher Education Area* (pp. 381-398). Springer.
- Colombo, O., & Boscia, T. (2022). Learners' Perceptions of Language Remote Teaching during the COVID-19 Pandemic. *Journal of Foreign Language Teaching and Learning*, 7(2), 245-268.
- Cocoradă, E., & Maican, M.A. (2021). Online Foreign Language Learning in Higher Education and Its Correlates during the COVID-19 Pandemic. *Sustainability*, 13(2), 1-21.
- Cutri, R. M., Mena, J., & Whiting, E. F. (2020). Faculty readiness for online crisis teaching: transitioning to online teaching during the COVID-19 pandemic. *European Journal of Teacher Education*, 43(4), 523-541. <https://doi.org/10.1080/02619768.2020.1815702>.
- Dresing, T., & Pehl, T. (2018). *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende*. Eigenverlag.
- Faria, A., Pereira-Neto, L., & Almeida, L. (2023). Dificuldades na adaptação acadêmica dos estudantes de ensino superior em tempos de pandemia. *Educação*, 48(1), 1-24. <https://doi.org/10.5902/1984644469514>
- Flick, U. (2011). *Qualitative Forschung. Eine Einführung*. Rowohlt.
- Flores, M. A., Simão, A. M., Barros, A., Flores, P., Pereira, D., Fernandes, E., Ferreira, P., & Costa, L. (2021). Ensino e aprendizagem à distância em tempos de COVID-19. Um estudo com alunos do Ensino Superior. *Revista Portuguesa de Pedagogia*, 55, 1-28.
- Kergel, D., & Heidkamp-Kergel, B. (2020). *E-learning, E-Didaktik und digitales Lernen. Diversität und Bildung im digitalen Zeitalter*. Springer VS. https://doi.org/10.1007/978-3-658-28277-6_3.
- Klimova, B. (2021). An Insight into Online Foreign Language Learning and Teaching in the Era of COVID-19 Pandemic. *Procedia Computer Science*, 192, 1787-1794. <https://doi.org/10.1016/j.procs.2021.08.183>
- Kuckartz, U., Dresing, T., Rädiker, S., & Stefer, C. (2008). *Qualitative Evaluation. Der Einstieg in die Praxis*. VS.
- Kuckartz, U. (2014). *Qualitative Text Analysis: A Guide to Methods, Practice and using Software*. Sage.
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Beltz Juventa.
- Leite, R., Silva, J., Santos, L., & Junior, R. (2021), O Uso das Tecnologias e Seus Desafios no Ensino Superior Remoto em Tempos de Pandemia de COVID-19, *Revista de Ensino Educação e Ciências Humanas*, 22(4), 517-525. <https://doi.org/10.17921/2447-8733.2021v22n4p517-525>.
- Mavridi, S. (2022). *Language Teaching Experiences During Covid-19*. British Council.

- Mayring, P., & Fenzl, T. (2014). Qualitative Inhaltsanalyse. In N. Baur, & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 543-556). Springer.
- Meyermann, A., & Porzelt, M. (2014). *Hinweise zur Anonymisierung qualitativer Daten*. Leibniz-Institut für Bildungsforschung und Bildungsinformation. <https://doi.org/10.25656/01:21968>.
- Mishra, L., Gupta, T., & Shree, A. (2020). Online Teaching-Learning in Higher Education during Lockdown Period of COVID-19 Pandemic. *International Journal of Educational Research Open*, 1, 1-8. <https://doi.org/10.1016/j.ijedro.2020.100012>
- Moreira, J. A., & Schlemmer, E. (2020). Por um novo conceito e paradigma de educação digital onlife. *Revista UFG*, 20(26), 2-35. <https://doi.org/10.5216/REVUFG.V20.63438>.
- Moore, M. G., & Kearsley, G. (2011). *Distance Education: A Systems View of Online Learning* (3rd ed.). Wadsworth.
- Patton, M. Q. (2015). *Qualitative Research and Evaluation Methods*. Sage, Thousand Oaks.
- Riemer, C. (2016). Befragung. In D. Caspari et al. (Eds.), *Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch*. Narr.
- Silverman, D. (2010). *Doing Qualitative Research. A Practical Handbook*. Sage.
- Petzold-Rudolph, K., Kay, K., Gericke, E., & Kuhlee, D. (2023). Digitale Lehr-Lernkonzepte zur Förderung digitaler LehrLernkompetenz. Ergebnisse aus dem Projekt DigiCom2, In R. Porsch, & C. Reintjes (Eds.), *Digitale Bildung im Lehramtsstudium während der Corona-Pandemie. Befunde, Erfahrungen, Perspektiven* (pp. 79-102). Waxmann. <https://doi.org/10.25656/01:28076>.
- Strömsdörfer, D. (Ed.) (2021). Herausforderung Digitalisierung - Lehr-Lern-Medien für DaF aus nationaler und internationaler Perspektive. Ergebnisse der digitalen Abschlusstagung der Germanistischen Institutspartnerschaft Freiburg-Belgorod am 06. November 2020, Pädagogische Hochschule Freiburg.
- Theelen, H., & van Breukelen, D. H. J.(2022). The didactic and pedagogical design of e-learning in higher education: A systematic literature review. *Journal of Computer Assisted Learning*, 38(5), 1286-1303. <https://doi.org/10.1111/jcal.12705>
- Toomnan, P. (2024). The Impact of Online Learning during the COVID-19 Pandemic on Language Learning Strategies, Proficiency, and Enjoyment among Thai EFL Students. *International Journal of Emerging Technologies in Learning*. 19(1),131-146. <https://doi.org/10.3991/ijet.v19i01.46401>.
- Torrissi-Steele, G. (2011). This thing called blended learning—A definition and planning approach. In K. Krause, M. Buckridge, C. Grimmer, & S. Purbrick-Illek (Eds.), *Research and development in higher education: Reshaping higher education* (Vol. 34, pp. 360-371).
- Traxler, P. (2023). Die Zukunft der Hochschullehre nach der Pandemie - Ergebnisse eines systematischen Literaturreviews zur postCOVID-19 Hochschullehre. *Workshop Gemeinschaften in Neuen Medien (GeNeMe)*. TUDpress - Verlag der Wissenschaften.
- Verbi Software (2019). *MAXQDA 2020 (Computer Software)*. Maxqda.com.

Os anos da mecanização (1940-1980): uma história da Tradução Automática

The years of mechanization (1940-1980): a history of Machine Translation

Ronaldo de Oliveira Batista

Universidade Presbiteriana Mackenzie (Brasil)

Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brasil)

ronaldo.obatista@gmail.com

Luciana Debonis

Universidade Presbiteriana Mackenzie (Brasil)

ludeb01@yahoo.com

Recebido: 04/06/2024

Aceite: 19/10/2024

Publicado: 23/12/2024

RESUMO: Narra-se uma história da Tradução Automática (TA) dos anos 1940-1980 a partir da Historiografia Linguística e do conceito de programas de investigação. Para tanto, duas perguntas são estabelecidas em função do período histórico selecionado: (1) Quais as características das ferramentas tecnológicas nesses anos?; (2) Quais impactos estão presentes nesses anos em relação à atuação do tradutor humano? O percurso metodológico adotado para apresentação de resultados neste artigo estabeleceu-se em três fases: 1ª. fase heurística: definição da periodização; seleção/tratamento do *corpus*; 2ª. fase hermenêutica: estabelecimento de diretrizes para a análise; 3ª. fase expositiva: apresentação de resultados em uma narrativa historiográfica. O material de análise foi organizado e hierarquizado em dois grandes conjuntos: documentos sobre funcionalidades e ferramentas; documentos sobre a história da TA. Argumenta-se, neste artigo, que em cerca de 40 anos a TA se caracterizou por iniciativas e empreendimentos de um programa de mecanização que começou a alterar o perfil do tradutor e sua inserção profissional.

PALAVRAS-CHAVE: Historiografia Linguística; Tradução Automática; Programas de Investigação; Mecanização.

ABSTRACT: We sketched the history of machine translation (MT) from the 1940s to the 1980s from a linguistic historiography perspective and the concept of research programs. For this purpose, two questions were established according to the selected historical period: (1) What are the characteristics and the use of technological tools? (2) What are the impacts such changes have on social contexts of use in relation to the performance of human translators? The methodological direction adopted to demonstrate the results in this article was established into three phases: 1st. heuristic phase: periodization definition; selection/treatment of corpus; 2nd. hermeneutic phase: guidelines establishment for the

analysis; 3rd. expository phase: results presentation through a historiographical narrative. The analysis material was organized and hierarchized into two large sets: documents regarding functionalities and tools and documents in connection with the history of MT. This article argues that for about 40 years, the MT was characterized by initiatives and operations of a mechanization program that began to change the translator's profile and professional insertion.

KEYWORDS: Linguistic Historiography; Machine Translation; Research Programs; Mechanization.

Introdução

Este artigo revê a história da Tradução Automática (TA) - “modalidade de tradução que consiste na automatização de uma parte ou do todo no processo de tradução de uma língua natural para outra” (Debonis 2021: 32) - entre 1940-1980 a partir da Historiografia Linguística, campo que estuda o desenvolvimento histórico do conhecimento sobre a linguagem, no qual estão também aqueles sobre tradução (cf. Altman, 2019; Batista, 2013, 2020; Koerner, 2014; Swiggers, 2019).

Duas perguntas nos guiam: (1) Quais as características das ferramentas tecnológicas nesses anos?; (2) Quais impactos estão presentes nesses anos em relação à atuação do tradutor humano? A argumentação principal é a de que em 40 anos a TA se caracterizou por empreendimentos de um “programa de mecanização”¹ que alterou o perfil do tradutor e sua inserção profissional.

As décadas de 1940-1980 na TA foram revistas por autores como Hutchins (2001), Poibeau (2017), Pym (1998). No entanto, o ponto de vista aqui adotado é distinto, pois insere a análise em um campo dos estudos linguísticos por meio de caracterização interna (características de ferramentas, métodos e visões teóricas subjacentes a estas) e externa (contextos científico-tecnológico e empresarial de produção, circulação e recepção de ferramentas e métodos) de práticas de tradução em torno do conceito de *programas de investigação*.

O percurso metodológico se deu em três fases e orientou a elaboração deste artigo: 1ª. fase, *heurística*: definição da periodização (1940-1980, os primeiros anos no século XX da história da TA - cf. Arnold *et al.*, 1994; Hutchins, 2001; Poibeau, 2017; Pym, 1998); seleção/tratamento do *corpus* (correspondências, relatórios,

¹ Este artigo retoma, reorganiza e sintetiza argumentações apresentadas em Debonis (2021).

experimentos, revisões históricas - Alpac, 1966; Bar-Hillel, 1951; Dostert, 1957; Holmes, 1972 [2000]; Hutchins, 2001; Masterman & Kay, 1960; Nelson, 1967; Oswald, 1952, 1953; Poibeau, 2017; Pym, 1998; Weaver, 1947-1949; Yngve, 1953); 2ª. fase, *hermenêutica*: com base no conceito teórico de *programas de investigação*, estabelecimento de perguntas (problemas a investigar, já citados) para apreensão de conteúdos presentes, pressupostos ou subentendidos no *corpus* com relação a seus diversos contextos de produção, circulação e recepção; 3ª. fase, *executiva*: apresentação desta narrativa histórica, caracterizada pelo formato expositivo em torno de causalidades e implicações explicativas de fatos diante de questionamentos incidentes sobre instâncias de produção, circulação e recepção de ideias e conteúdos examinados na fase hermenêutica.

Como apontado, a história da TA é observada pelo conceito de *programas de investigação*: modos de tratamento da linguagem definidos por procedimentos adotados e resultados alcançados, derivados de concepções teóricas. O conceito é de Swiggers (1981). Seus programas de investigação (descritivista, de correspondência, sociocultural, de projeção), originalmente definidos para teorias linguísticas, precisam constantemente passar por revisões diante de novas tendências e confluências na linguística e nas práticas a ela relacionadas (como a tradução).

Este artigo, ao rever Swiggers (1981), propõe a existência na história da TA de um “programa de mecanização”. Insere, assim, em outras narrativas históricas sobre métodos de tradução (ampliando o escopo destas), uma perspectiva que articula características técnicas e científicas de um modelo tradutório a elementos sociais que sustentaram práticas profissionais de tradutores.

Para tanto, apresenta uma narrativa histórica em três seções principais (além de uma Introdução na qual o problema central é colocado e de uma Conclusão que retoma os questionamentos iniciais e os responde): a) a primeira, que trata de um clima intelectual e científico nos anos 1940-1970; b) a segunda, que propõe a caracterização de um programa de mecanização na história da TA; c) a terceira, que observa aspectos sociais do impacto desse período de mecanização na história da TA.

1. Como cenário para a TA, um clima intelectual formalista

Estudos linguísticos dos anos 1940-1970 foram em grande parte estruturalistas de feição formalista. Um cenário teórico-metodológico que seria alterado durante os anos 1970 com a presença de teorias interdisciplinares (como a sociolinguística e a psicolinguística), teorias do texto e do discurso e teorias funcionalistas.

Neste artigo, *estruturalismo* é denominação abrangente e se refere a propostas teórico-metodológicas em período que vai da década de 1920 até finais da década de 1970. Dentre estas: o descritivismo nos Estados Unidos (Leonard Bloomfield [1887-1949] e Kenneth Pike [1912-2000]); estruturalismos derivados do Círculo Linguístico de Praga (1920-1930); a glossemática de Louis Hjelmslev (1899-1965); a semântica estrutural de Algirdas Julien Greimas (1917-1992); o gerativismo de Noam Chomsky nas teorias padrão e padrão estendido. A definição *estruturalista* refere-se antes de tudo a um ponto de vista, mais do que a uma corrente teórica única: é uma concepção geral de língua como conjunto organizado e regado de unidades em permanente relação. Já *formalista/formalismo* denotam abordagem imanente de aspectos linguísticos. Para os formalistas, atribuem-se: preexistência, autonomia e primazia analítica da forma; forma anterior ao uso; representações visuais correlacionadas a regras estruturais ou aspectos mentais; explicações internas ao sistema (sem apelo ao social ou à subjetividade dos falantes).

Algumas dessas correntes formalistas influenciaram, por hipótese, procedimentos de tradução, mesmo que não reconhecidas explicitamente fora do universo acadêmico em humanidades. Um paradigma formalista estruturalista que não foi homogêneo em teorias e métodos, pois diferentes propostas foram divulgadas e defendidas. Algumas tiveram seus momentos de destaque e outras de ostracismo. Diante desse pressuposto, diferentes modos de conceber tarefas do linguista eram possíveis.

Nesses anos 1940-1970, teorias formalistas estruturalistas analisavam os sistemas linguísticos como fim em si mesmos. E a forma linguística era definida por relações estruturais e regulares entre as unidades de um sistema. Nessa vertente, teríamos os “linguistas-engenheiros” (Thomas, 2020), que dariam destaque para:

representações formais correlacionadas a regras; explicações internas ao sistema (sem nenhum apelo ao social ou à subjetividade dos falantes); generalizações; observação (com base em método dedutivo) de fatos gramaticais específicos.

Seria essa linguística (em toda sua diversidade) que comporia cenário de fundo para propostas de tradução (especialmente as de natureza técnica) com uso de ferramentas e programas computacionais que mecanicamente fariam transições interlinguísticas.

2. Um programa de mecanização na história da TA

Neste artigo, argumenta-se que os primeiros anos da história da TA foram anos de um “programa de mecanização”, com pesquisas em protótipos de sistemas computacionais motivadas por diferentes conflitos entre povos e nações, diante de demandas de vigilância e compreensão rápida de línguas de outros povos e países.

E por que a caracterização como programa de mecanização? Seguindo diretrizes classificatórias de Swiggers (1981) para outro objeto analítico e contexto científico, destacam-se como elementos para existência desse programa na história da TA:

QUADRO 1 - Caracterização do programa de mecanização na TA

CONCEPÇÃO TEÓRICA: visão de língua como sistema codificado (criptografia) e de gramática como estrutura de relações formais regulares entre unidades.
MÉTODO: técnica de tradução no nível da palavra (incidência no léxico) e modelo de tradução com base em dicionário.

(Fonte: elaborado pelos autores)

Situar mudanças no percurso histórico da TA é também propor paralelo com marcos nos estudos linguísticos em termos de alterações em paradigmas teóricos e metodológicos que delimitaram objetivos e alcances de práticas de pesquisa em torno da descrição e análise linguísticas.

O programa de mecanização está relacionado, até pelo recorte temporal e pelo espaço geográfico no qual inicialmente estava circunscrito (os EUA), ao descritivismo praticado por linguistas formalistas norte-americanos (cf. Altman,

2021) como Leonard Bloomfield e seus seguidores (os pós-bloomfieldianos ou distribucionalistas).

O que permite essa vinculação é em essência o teor mecanicista da linguística de extração bloomfieldiana. Ao privilegiar descrições sincrônicas dos níveis fonológico e morfológico, por meio da seleção de unidades e da explicação dos seus modos de combinação e atuação no sistema, o descritivismo norte-americano (como é reconhecido o paradigma com origem em Bloomfield e Kenneth Pike) associou-se a uma concepção behaviorista de linguagem. Daí, portanto, a aproximação entre o programa de mecanização aqui exposto e o descritivismo de base behaviorista.

O programa de mecanização agrupa pesquisas com objetivo de mecanizar a gramática, ou seja, esforços voltavam-se para criação de sistemas em que a máquina aplicasse regras gramaticais das línguas naturais. Buscava-se programar a máquina para realizar tarefas de tradução como se fosse uma operação matemática. Nesta apenas a aplicação da regra certa poderia transferir um texto de partida para uma língua de chegada (LC), sem considerar o que aconteceria no caminho da transferência. Primeiras ferramentas usavam o modelo de “tradução palavra por palavra” (WMT), no qual eram usados dicionários e glossários bilíngues para transferência linguística.

Pesquisadores tentaram simplificar dicionários bilíngues, fornecendo equivalentes únicos para palavras da língua de partida (LP) que abrangessem uma ampla maioria de sentidos. Assim, os experimentos buscavam, na sua maioria, desenvolver regras de programação para traduzir de forma direta, de uma LP para uma LC específica apenas, com mínima análise e reorganização sintática, evitando análise de contextos e mantendo a ordem das palavras da LP.

Alguns grupos de pesquisa investigaram, ainda, possibilidades de representações “interlínguas”, ou seja, códigos ou símbolos independentes tanto na LP como na LC, em que a tradução ocorreria em duas etapas: 1) a partir da LP para a “interlíngua” e 2) a partir da “interlíngua” para a LC. Porém, essa era uma visão de pesquisa mais a longo prazo e a maioria se dedicou a pesquisas linguísticas básicas, isso porque alguns pesquisadores já acreditavam que apenas

investigações mais específicas sobre o processo de pensamento humano (posteriormente chamado de “inteligência artificial” ou “ciência cognitiva”) poderia melhorar a qualidade da TA (Hutchins, 2001: 11).

A base teórica era o conceito da tradução direta. Primeiros resultados não foram satisfatórios, com baixa qualidade na tradução. Apesar da qualidade ruim no texto da LC, pesquisas e usos da ferramenta continuaram. Isso atendia à agilidade e à redução de custos em comparação com o tradutor humano.

Esse processo de uma tradução direta, ancorada na possibilidade de criação de máquinas, tem em essência um vínculo, mesmo que muitas vezes não intencional, com a visão mecanicista de língua que caracterizou parte do estruturalismo linguístico.

A construção de máquinas para TA com base puramente mecânica estendeu-se até a década de 1960, mas recursos computacionais disponíveis representavam barreira para evolução dos modelos tecnológicos. Nesse período, a máquina só executava a tradução em uma direção, por exemplo, do russo para o inglês, e não conseguia realizar o caminho contrário, para isso era preciso outra máquina.

Primeiras ferramentas e protótipos de TA tiveram menor envolvimento de linguistas e maior concentração de especialistas da área de exatas e da computação (que estava começando). O principal envolvimento de linguistas se dava na interpretação gramatical, para estabelecer regras gramaticais a serem aplicadas na máquina mecanicamente, assim como, na pós-edição do texto traduzido pela máquina, tarefa essa desempenhada por linguistas e tradutores profissionais.

Os grupos de pesquisa atuantes nesse programa estavam em grande parte nos EUA e se formaram, principalmente, dentro de universidades com investimento público e privado. Esses grupos contribuíram para a produção acadêmica nas décadas de 1950-1960, impulsionando avanços computacionais, motivados pela demanda da TA e que serviram, posteriormente, como fundamento para o desenvolvimento de linguagens de programação.

QUADRO 2 - Grupos de pesquisa norte-americanos em 1950-1960

Grupo de Erwin Reiffler, Universidade de Washington (Seattle, EUA): otimizou a abordagem da tradução direta com base em dicionário que envolveu a construção de grandes dicionários bilíngues em que a informação lexical era usada também para resolver problemas gramaticais, sem o uso de análise sintática.
Grupo de Leon Dostert, Universidade de Georgetown (Washington, EUA): conduziu análises de textos apenas quando as informações da gramática tradicional eram inadequadas. Desenvolveu o sistema <i>Georgetown Automatic Translation</i> (GAT), que possuía três níveis de análise: morfológica, sintagmática e sintática. O sistema foi instalado na Itália em 1963 pela Euratom e posteriormente adotado pela Comissão de Energia Atômica Americana em 1964.
Grupo de Anthony Oettinger, Universidade de Harvard (Massachusetts, EUA): concentrou as pesquisas na compilação de um dicionário russo-inglês e em 1959 avançou no desenvolvimento de sistema para identificação de sequências de categorias gramaticais e sua predição probabilística.
Grupo de Victor Yngve, MIT (Massachusetts, EUA): em 1953 concentrava estudos na relação sintática. A LP era analisada como representações de estruturas de frase, em que uma rotina de transferência de estrutura convertia em estruturas de frases equivalentes na LC. O principal problema encontrado era a barreira semântica do texto traduzido. Chomsky fez parte desse grupo por pouco tempo, mas não se encontram evidências de abordagem chomskiana nas pesquisas em TA nesta época.
Grupo de Winfried Lehmann, Universidade do Texas (Austin, EUA): grupo atuando no <i>Linguistic Research Center</i> em 1958, que concentrava estudos na relação sintática. Tentou construir sistema que pudesse fazer tradução bidirecional com base na abordagem da transferência sintática, deixando fundamentos para o posterior desenvolvimento do sistema METAL.
Grupo de Sydney Lamb, Universidade da Califórnia (Berkeley, USA): concentrou-se no desenvolvimento de rotinas de dicionários eficientes e de uma teoria linguística adequada para TA. Este conceito seguia a ideia de relações paralelas com a arquitetura de computadores (nós e redes), em que a tradução era vista como processo de codificação e decodificação.

(Fonte: elaborado pelos autores)

QUADRO 3 - Grupos de pesquisa fora dos EUA em 1950-1960

União Soviética (atual Rússia): concentravam-se em sistemas inglês-russo e na análise morfológica e sintática para a transformação da LP por meio de regras gramaticais. Apesar de o inglês ser o principal idioma estudado devido questões políticas entre os países, as pesquisas ampliaram para uma diversidade maior de idiomas do que nos EUA. Os principais grupos de pesquisa foram: o grupo de D. Y. Panov no <i>Institute of Precision Mechanic</i> (hoje <i>Lebedev Institute of Precision Mechanics and Computer Engineering</i>), o grupo de Ljapunov, Kulagina e Mel'cuk no <i>Steklov Mathematical Institute</i> e o grupo de Nikolaj Andreev na Universidade Estadual de Leningrado.
Londres, Reino Unido: o grupo de Margaret Masterman na <i>Language Research Unit</i> de Cambridge adotou a linha de abordagem "interlíngua" que consiste no método de transformação do texto fonte numa representação abstrata de língua independente (a interlíngua) e, em seguida, o texto de chegada é gerado a partir da interlíngua. No caso das pesquisas de Cambridge, o desenvolvimento do método interlíngua foi baseado em palavra por palavra produzindo traduções denominadas "pidgin".
Milão, Itália: o grupo de Silvio Ceccato na Universidade de Milão também concentrou suas pesquisas no método da interlíngua e o desenvolveu com base em processos cognitivos, especialmente no conceito de análise de palavras e suas possíveis correlações com outras palavras no texto.

(Fonte: elaborado pelos autores)

Outros grupos fora dos EUA avançaram com protótipos de sistemas de TA nas décadas seguintes. Como é o caso do grupo TAUM liderado por John Chandioux na Universidade de Montreal, no Canadá, responsável pelo desenvolvimento do sistema Meteo adotado em 1977 pelo Centro de Meteorologia Canadense. Além disso, grupos de pesquisa no território europeu e no Japão deram seguimento ao desenvolvimento de ferramentas de TA. Dentre estes: o grupo que desenvolveu o sistema GETA-Ariane na França; o MU no Japão; o SUSY na Alemanha; o AMPAR (inglês-russo) e NERPA (alemão-russo) na ex-União Soviética.

Em 1970, na Universidade de Montreal, a pesquisa do grupo TAUM sobre o sistema Meteo (para tradução inglês-francês de previsões meteorológicas) aplicava vocabulário e sintaxe específicos dos relatórios meteorológicos. O sistema Meteo, baseado em tradução direta, empregava dois dicionários bilíngues para processamento da TA. Sendo, a) um de expressões meteorológicas, incluindo por volta de 300 verbetes e b) outro com categorizações lexicais, ou seja, para cada par inglês-francês teria sua correspondência semântica e variações morfológicas; este incluía por volta de 1.200 verbetes. O processamento da TA no Meteo era feito por meio da aplicação de um *parser* (analisador) capaz de analisar uma sequência de frases na LP e buscar sua correspondência na LC, de acordo com as regras especificadas pelos dicionários aplicados.

Processamento de TA a partir de tradução direta, característico do programa de mecanização, pode ser observado também no desenvolvimento de sistemas como Metal (1970) - parceria entre a Universidade do Texas (EUA) e a empresa Siemens (Alemanha) - para tradução do inglês-alemão, e Sysgram (1976) - desenvolvido pela empresa norte-americana Sysgram - para tradução de diferentes pares de idiomas, que consistia na tradução das sentenças analisadas na LP em suas "formas normais", ou seja, sem elementos lexicais interlínguas. Porém, problemas nos níveis de análise detectados eram difíceis de serem corrigidos.

A importância desses grupos é fundamental. Foram as elites universitárias responsáveis pela difusão de uma teoria e legitimação no ambiente acadêmico. A

rede de influências foi consolidada por esses grupos, que promoveram circulação da produção acadêmica e aplicação profissional de suas propostas.

A divulgação do trabalho desses grupos avançou, e em 1954 a primeira revista sobre TA intitulada "*Mechanical Translation*" foi lançada por William Locke e Victor Yngve, que sucederam Bar-Hillel no Instituto de Tecnologia de Massachusetts (MIT) a partir de 1953. Em 1954, foi concluída a primeira tese de doutorado em TA intitulada "*A study for the design of an automatic dictionary*", de Anthony G. Oettinger (Harvard), que estudou um dicionário mecânico em russo que pudesse ser atualizado automaticamente. No ano seguinte, em 1955, o primeiro livro sobre TA intitulado "*Machine Translation of Languages: fourteen essays*" foi lançado e era uma coleção editada por William Locke e Andrew Booth, incluindo artigos acadêmicos publicados por Yehoshua Bar-Hillel (MIT), Leon Dostert (Georgetown), Anthony Oettinger (Harvard), Erwin Reifler (Universidade de Washington) e Victor Yngve (MIT), além do memorando de Warren Weaver de 1947.

Apesar de várias iniciativas para a divulgação da TA, as pesquisas publicadas ainda eram poucas. Não encontramos trabalhos na área das linguagens nessa época que discutam TA. A produção era fundamentalmente sobre estudos técnicos, sem interesse para as humanidades na discussão sobre impactos sociais da revolução tecnológica que iniciativas como as de TA provocariam.

A publicação de um relatório em 1966 criticando investimentos em TA provocou desmonte dos grupos norte-americanos. Redigido pela ALPAC, resultou no redirecionamento do apoio financeiro para produção científica voltada a aparatos bélicos e às ciências espaciais (os EUA passavam pela Guerra do Vietnã [1955-1975] e começavam sua Corrida Espacial [1960-1975]).

Os efeitos do relatório também foram percebidos em grupos fora do território norte-americano, mas estes conseguiram atingir resultados importantes, o que motivou uma retomada gradativa de grupos de pesquisa nos EUA a partir da década de 1980.

A década de 1980 também foi marcada pelo estouro nas pesquisas em tecnologia no Japão e, conseqüentemente, aumentou o interesse desse país no desenvolvimento de sistemas de TA. Essas pesquisas também influenciaram o

interesse pela TA de outros países daquele continente como Coréia, Taiwan, China e Malásia.

Nesse ponto, os grupos começaram a se interessar pelo processo do pensamento humano (o que futuramente seriam os estudos em IA) no PLN e acreditavam que naquele caminho encontrariam as soluções para os problemas da TA. Esses grupos impulsionaram a consolidação da TA como uma área de pesquisa importante. No entanto, ainda se tratava de um campo de pesquisa complexo, tanto do ponto de vista técnico como de gestão organizacional e de infraestrutura, e caro em relação ao esforço de formação de equipes e tempo para obter resultados.

Influenciado por projetos globais, o grupo de pesquisa GETA, em Grenoble na França, começou a desenvolver o sistema Ariane nos anos 1980. O sistema destacou-se pela flexibilidade, modularidade e algoritmos avançados para manipulação de estruturas de árvores. O grupo introduziu conceitos de gramáticas estáticas e dinâmicas, permitindo diferentes níveis de representação em estruturas de árvore rotuladas, o que proporcionou flexibilidade em transferências multiníveis.

O mercado de TA teve uma expansão e vários sistemas entraram em funcionamento em instituições públicas e grandes corporações privadas. Multinacionais investiram no desenvolvimento de seus próprios sistemas de TA e, posteriormente, empresas de *softwares* entraram no mercado oferecendo opções personalizadas também. Alguns desses primeiros sistemas serviram de base para o desenvolvimento de programas futuros.

QUADRO 4 - Primeiros sistemas comercializados nos anos de 1980

SISTEMA	GRUPO DE PESQUISA
GAT	Universidade de Georgetown (USA)
METEO	Universidade de Montreal (Canadá)
TITUS	<i>Institut Textile de France</i> (França)
CULT	Universidade de Hong Kong (China)
Systran	Systran (USA)
Logos	Logos Corporation (USA)
METAL	Universidade do Texas em parceria com a empresa Siemens (USA e Alemanha)

(Fonte: elaborado pelos autores)

Alguns projetos deixaram importantes legados, sendo precursores da retomada dos estudos em TA nos anos de 1980, pavimentando vias para a grande virada na área a partir da década seguinte. Como exemplo, está o grupo GETA (e seu sistema Ariane), que desenvolveu o projeto Eurotra, em que interações entre a TA e a linguística computacional começaram a ser mais efetivos

Nesse período, pesquisas em TA impulsionaram diálogos com áreas na linguística formalista (semântica formal, linguística matemática, coleta de dados quantitativos) e na linguística computacional.

No Japão, por exemplo, o sistema MU desenvolvido na Universidade de Kyoto por Makoto Nagao, seguindo modelo semelhante ao GETA-Ariane, destacou-se pelo uso da análise gramatical, de representações de árvores de dependência (representação visual empregada em teorias linguísticas formalistas) e desenvolvimento de um ambiente de programação para escrita gramatical. Sua pesquisa influenciou vários projetos de TA no Japão, incluindo sistemas comerciais.

Já a pesquisa conduzida na Universidade de Saarland, em Saarbrücken na Alemanha, desenvolveu o sistema SUSY que empregava diferentes técnicas, como regras de estrutura frasal, transformações gramaticais e análise de casos (elementos presentes em linguísticas formalistas como a gramática gerativa). Seu foco principal era o tratamento detalhado de línguas como russo e alemão, mas também investigou inglês e francês.

O programa de mecanização ficou marcado com grande importância para diferentes campos no que diz respeito à aplicação de computadores para realização de tarefas "intelectuais" humanas. A demanda pela TA nesse período significou uma das grandes motivações para pesquisas em computação, servindo como fundamento para o desenvolvimento de linguagens de programação.

3. O tradutor humano profissional durante o programa de mecanização

A necessidade de atender a volume crescente de traduções que não poderia ser resolvido apenas com tradução humana, bem como a demanda por agilidade e economia de custos, levou ao surgimento do programa de mecanização na TA.

Primeiros protótipos de sistemas de TA foram testados na tradução de artigos científicos, documentos que tinham grande volume impresso e exigiam tradução rápida. Com isso, a TA começou a ser idealizada porque, apesar de entregar um texto de pouca qualidade, baixa fluência e no formato de tradução palavra por palavra, ainda o fazia de maneira mais rápida em comparação ao trabalho de um tradutor humano. Consequentemente, aceitava-se uma tradução de menor qualidade, visto que se buscava entendimento geral do assunto para manter, em contexto de conflitos políticos, por exemplo, a observação dos movimentos do inimigo.

Com novas possibilidades de automatização da tradução, observa-se nova função atribuída aos tradutores: o trabalho de pós-edição da tradução gerada pela máquina. Essa função ainda é bastante executada por tradutores, revisores, linguistas, entre outros, atualmente, mesmo com uma TA de qualidade bem superior às geradas durante o programa de mecanização.

A perspectiva de idealização de uma TA perfeita (conceito de FAHQMT [*Fully Automatic High Quality Machine Translation*]) defendida por Bar-Hillel (1951) afirmava que a necessidade de pré/pós-edição realizada por um tradutor humano no processo não representaria uma TA. Essa retórica desmoralizou a TA e o trabalho de pesquisa desenvolvido naquele período, ocasionando a suspensão dos estudos por uma década, em especial nos EUA.

O programa de mecanização na TA, tal como aqui proposto, não apresentou especificamente base teórica articulada em um grupo teórico em diálogo com estudos de tradução (a rigor, estudos dessa área se institucionalizaram na década de 1970). Foi escasso o envolvimento de profissionais da linguagem (tradutores e linguistas) no desenvolvimento de ferramentas, que ficaram a cargo de especialistas das ciências exatas e da ciência da computação.

Já nos anos 1980, tradutores começaram a reconhecer, em alguns tipos de trabalho, as vantagens dos programas de computação no tratamento de dados e de dicionários. No entanto, um avanço mais significativo com engajamento de profissionais da linguagem em diálogo com aqueles já envolvidos no programa de mecanização se daria apenas nos anos 1990.

Como resultado social da presença de ferramentas computacionais durante os anos 1940-1980, o processo de tradução como atividade intelectual elitizada, gradativamente, começou a se transformar em conhecimento de domínio comum ao longo dos anos seguintes. Essa alteração social exigiu cada vez mais do tradutor o desenvolvimento de habilidades em novas tecnologias para atender a uma demanda que despontaria nas próximas décadas no mercado para serviços linguísticos em TA.

Colocar em discussão o trabalho do tradutor profissional, como proposto neste artigo em visão historiográfica, auxilia-nos na compreensão de como a relação tradutor e ferramentas foi sendo moldada em diferentes tempos históricos nos quais a tecnologia possibilitou novos horizontes e novos mecanismos. O passado, nesse sentido, é meio de compreensão de nosso presente. Na nossa época, as tecnologias de informação e comunicação passam por inovações constantes e necessitam ser objeto de reflexão a respeito, mais uma vez na história, da relação homens e máquinas.

Considerações finais

Durante os anos 1940-1980, a TA foi concebida e caracterizou-se pelo desenvolvimento de ferramentas tecnológicas pensadas para mecanização da tradução com base no processo de transformação de estruturas lógicas. Empregava-se o conceito de tradução direta, executada com base em regras gramaticais e técnica de tradução palavra por palavra. Esse processo de tradução era processado a partir do desenvolvimento de dicionários, com apresentação de terminologia específica para área e equivalentes únicos na língua de chegada que tentavam abranger ampla variedade de sentidos da língua de partida.

Num cenário intelectual de estudos linguísticos predominantemente estruturalista e formalista, grupos de pesquisa ao redor do mundo se concentraram no desenvolvimento de ferramentas de TA a partir da perspectiva das relações estruturais e regulares da língua, com técnicas para codificação e aplicação de regras gramaticais formais.

Uma necessidade de mecanização da tradução e, conseqüentemente, uso da TA nesse período buscava atender à demanda por rapidez na tradução de grandes volumes de textos, especialmente, por se tratar de contexto de guerra (Guerra Fria [1947-1991]), em que a decodificação rápida de informações e vigilância de um potencial inimigo eram cruciais.

Com isso, a baixa qualidade da tradução entregue pela máquina era compensada pela velocidade e economia de custos que proporcionava, visto que o objetivo era obter um entendimento geral de textos, em sua maioria técnico-científicos.

O processo de TA introduziu a necessidade de pós-edição de traduções geradas pela máquina, função desempenhada por tradutores humanos, na época, e ainda demandada para certos tipos de conteúdo atuais. Com a inserção do trabalho realizado pelo tradutor humano num contexto de uso de ferramentas tecnológicas, a partir dos anos 1980, presenciamos o desenvolvimento de ferramentas como dicionários personalizados, sistemas de gerenciamento de texto, memórias de tradução que facilitaram o trabalho de tradutores humanos, melhorando eficiência e consistência no trabalho.

A necessidade de mecanização da tradução e o surgimento da TA foram motivadores para desenvolvimentos futuros de estudos computacionais, com avanços significativos em linguagens de programação, consolidando bases para evolução tecnológica de diferentes áreas de estudos nos anos seguintes.

Referências

- ALPAC. (1966). *Language and machines: computers in translation and linguistics*. Division of Behavioral Sciences, National Academy of Sciences, National Research Council.
- Altman, C. (2021). *A guerra fria estruturalista: estudos em historiografia linguística brasileira*. Parábola.
- Altman, C. (2019). História, estórias e historiografia da linguística brasileira. In R. de O. Batista (Org.). *Historiografia da Linguística* (pp. 19-44). Contexto.
- Arnold, D., Balkan, L., Meijer, S., Humphreis, R. L., & Sadler, L. (1994). *Machine Translation: an introductory guide*. Blackwell.
- Bar-Hillel, Y. (1951). The present state of research on mechanical translation. *American Documentation*, 2(1), 153-165.
- Batista, R. de O. (2013). *Introdução à historiografia da linguística*. Contexto.

- Batista, R. de O. (2020). *Fundamentos da pesquisa em historiografia da linguística*. Mackenzie.
- Debonis, L. (2021). *Um estudo historiográfico da evolução da tradução automática* [Dissertação de Mestrado, Universidade Presbiteriana Mackenzie]. Adelpha Repositório Digital. <https://dspace.mackenzie.br/bitstreams/2b41d2ab-3122-4569-8111-a3286e28784e/download>
- Dostert, L. (1957). Systems of logic in machine translation. *Eighth Annual Round Table Meeting on Linguistics and Language Studies*. Washington D.
- Holmes, J. (2000). The name and nature of translation studies. [republicação de texto original de 1972] In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 172-185). Routledge.
- Hutchins, W. J. (2001). Machine translation over fifty years. *Histoire Épistémologie Language*, 23(1), 7-31.
- Koerner, K. (2014). *Quatro décadas de historiografia linguística: estudos selecionados*. Centro de Estudos em Letras - Universidade de Trás-os-Montes e Alto Douro.
- Masterman, M., & Kay, M. (1960). *Mechanical pidgin translation: an estimate of the research value of "word-for-word" translation into a pidgin language, rather than into the full normal form of an output language*. Cambridge Language Research Unit.
- Nelson, B. (1967). Machine Translation: Committe Skeptical over Research Support. *Science*, 155(1), 1-2.
- Oswald, V. (1952). Word-by-word translation. *Conference on Mechanical Translation, Massachusetts Institute of Technology*. Massachusetts Institute of Technology.
- Oswald, V. (1953). An idioglossary for mechanical translation. *Modern Language Forum*, 38, 3-4.
- Poibeau, T. (2017). *Machine translation*. The MIT Press.
- Pym, A. (1998). *Method in translation history* (2nd. ed.). Routledge.
- Swiggers, P. (1981). The history writing of linguistics: a methodological note. *General Linguistics*, 21(1), 11-16.
- Swiggers, P. (2019). Historiografia da Linguística: princípios, perspectiva, problemas. In R. de O. Batista (Org.), *Historiografia da Linguística* (pp. 45-80). Contexto.
- Thomas, M. (2020). *Formalism and functionalism in linguistics*. Routledge.
- Weaver, W. (1947). *Correspondência*. The Rockefeller Foundation.
- Weaver, W. (1948). *Correspondência*. The Rockefeller Foundation.
- Weaver, W. (1949). *Correspondência*. The Rockefeller Foundation.
- Yngve, V. H. (1953). *Mechanical Translation. Quarterly Progress Report of the Research Laboratory of Electronics*. Massachusetts Institute of Technology.

Preposição: uma categoria intermediária no continuum léxico > gramática¹

*Preposition: an intermediate category in the continuum
lexicon > grammar*

Sueli Maria Coelho

Faculdade de Letras da Universidade Federal de Minas Gerais (Brasil)
sucoelho@ufmg.br

Recebido: 15/02/2024

Aceite: 27/05/2024

Publicado: 23/12/2024

ABSTRACT: From a cognitivist and multisystemic perspective on the conception of language, this article revisits the class of prepositions in Portuguese, re-discussing their concept and their traditional categorization in terms of form and function. The core of the discussion consists of a proposal to categorize the eleven most productive essential prepositions in Portuguese according to the prototype of the items, defined based on a matrix of features that combines categorical criteria (degree of grammaticalization), syntactic (transitivity potential) and cognitive (representation concrete and metaphorical space). In this analysis methodology, the most prototypical examples of the class are those that combine the greatest number of features in the matrix, while the most peripheral are those that present fewer features that define the category. The results obtained indicate a specialization in the use of the most peripheral exemplars in the affix function.

KEYWORDS: Preposition; Cognitive Linguistics; Categorization; Portuguese.

RESUMO: A partir de uma perspectiva cognitivista e multissistêmica de concepção de língua(gem), neste artigo revisita-se a classe das preposições no português, discutindo seu conceito e sua tradicional categorização quanto à forma e quanto à função. O cerne da discussão consiste numa proposição de categorização das onze preposições essenciais mais produtivas do português segundo a prototípia dos itens, definida com base numa matriz de traços que combina critérios categoriais (grau de gramaticalização), sintáticos (potencial de transitividade) e cognitivos (representação espacial concreta e metafórica). Nessa metodologia de análise, os exemplares mais prototípicos da classe são aqueles que combinam o maior número de traços na matriz, ao passo que os mais periféricos são aqueles que apresentam menos traços definidores da categoria. Os resultados obtidos sinalizam para uma especialização de uso dos exemplares mais periféricos na função de afixo.

PALAVRAS-CHAVE: Preposição; Linguística Cognitiva; Categorização; Português.

¹ A pesquisa que deu origem a este artigo resulta de um estudo mais amplo sobre a classe das preposições desenvolvido durante um estágio de pós-doutoramento na Universidade Estadual de Campinas, sob a supervisão do Prof. Ataliba Teixeira de Castilho.

Introdução

Neste texto, propomo-nos a refletir acerca de uma classe de palavras cuja descrição se fez de modo bastante sucinto na tradição gramatical, mas que, no âmbito das pesquisas linguísticas, tem tido sua importância funcional desvelada e atraído os olhares de vários pesquisadores, em diversas perspectivas teóricas. Trata-se da *classe das preposições*, que será aqui abordada numa concepção cognitivista e multissistêmica de língua(gem). Segundo a abordagem cognitivista, “os significados codificados na linguagem se relacionam e refletem nosso sistema conceitual, que constitui nossa ‘representação’ da realidade” (Tyler e Evans, 2003: X, tradução nossa²). Tal como defendido por Fauconnier (1997), por Jackendoff (1983, 1987, 1990, 1992) e por Langacker (1987, 1991), nossa percepção da realidade é construída com base nas nossas experiências na interação com o mundo físico-espacial que habitamos e mediada por nosso sistema perceptivo. Sob esse ponto de vista, busca-se uma visão integradora do fenômeno da linguagem segundo a qual a mente “não é separada do corpo. Ao contrário, o pensamento é *corporificado* no sentido de que a sua estrutura e a sua organização estão diretamente associadas à estrutura de nosso corpo, bem como às nossas restrições de percepção e de movimento no espaço” (Martelotta e Palomanes, 2009: 181). Esse entendimento conduz à máxima de que o espaço é o domínio básico da experiência humana, sendo, portanto, fonte para entender os demais domínios abstratos que dele se originam por meio de processos de extensão metafórica. Nesse viés, o significado “não tem nada a ver com a relação de pareamento entre linguagem e mundo. [...] A significação linguística emerge de nossas significações corpóreas, dos movimentos de nossos corpos em interação com o meio que nos circunda” (Oliveira, 2001: 34) e se amplia a partir de operações cognitivas de transferência de significados entre domínios de abstração gradativa. Em face disso, a perspectiva cognitivista mostra-se bastante apropriada para estudar as preposições, já que “seu sentido de base é reconhecível quando elas expressam as

² Do original: “[...] the meanings encoded in language relate to and reflect our conceptual system, which constitutes our ‘representation’ of reality.” (Tyler e Evans, 2003: X)

categorias POSIÇÃO NO ESPAÇO, DESLOCAMENTO NO ESPAÇO e DISTÂNCIA NO ESPAÇO” (Castilho, 2010: 585).

Justificada a opção pelo quadro teórico cognitivista para abordar o tema de que nos ocupamos neste artigo, resta-nos fazê-lo também no que tange à concepção multissistêmica de língua(gem). Estamos assumindo, com base em Castilho (2007, 2009a, 2009b, 2010), que a língua é um sistema dinâmico e, como tal, complexo; “um conjunto de categorias agrupadas simultaneamente em quatro subsistemas: (1) Léxico, (2) Discurso, (3) Semântica, (4) Gramática” (Castilho, 2009: 129, tradução nossa³), os quais operam autonomamente no sentido de que não se pressupõe nenhuma hierarquia entre eles, o que implica admitir, portanto, que não há subsistema central nem periférico. Sob esse prisma, o que é “central na língua é o *dispositivo sociocognitivo* que rege os sistemas, fazendo de cada um deles uma forma de representação das categorias cognitivas” (Castilho, 2010: 83, itálicos do autor). Considerando-se a inquestionável dinamicidade do sistema linguístico como um todo, há de se admitir a inerência da mudança, que afeta igualmente cada um desses quatro subsistemas por meio de seus respectivos processos de inovação: lexicalização, discursivização, semantização e gramaticalização. Em se tratando da classe das preposições, os processos mais produtivos são a semantização e a gramaticalização, motivo pelo qual serão referidos ao longo deste texto quando se fizer pertinente.

Como já anunciado, nosso objetivo geral nos domínios deste artigo é trazer à baila reflexões acerca da classe das preposições, buscando transcender a noção de que tais palavras invariáveis constituem uma classe fechada cuja função é meramente relacional. No bojo desse propósito mais amplo, intentamos apresentar uma categorização original para a classe das preposições no quadro da Linguística Cognitiva, propondo uma matriz de traços conceptuais e estruturais definidores da categoria, a partir da qual buscamos estabelecer os exemplares mais prototípicos do conjunto. Iniciaremos nossa tarefa pela conceituação de nosso objeto de estudo, bem como por sua tradicional categorização (cf. seção 1). A partir da análise das limitações que tal categorização impõe à classe, propomos, na segunda seção, uma

³ Do original: “[Language-as-a-product] is a set of categories grouped together at the same time in four subsystems: (1) Lexicon (2) Discourse (3) Semantics (4) Grammar” (Castilho, 2009: 129).

matriz de traços que combina critérios categoriais, sintáticos e cognitivos, com vistas a delinear efeitos de prototipicidade no inventário das preposições. Isso posto, passamos à síntese de nossas reflexões, apontando para perspectivas futuras de investigação sobre o tema, sobretudo à luz dos resultados alcançados.

1. Preposição: conceito e categorização

O termo *aposição* refere-se a um processo de combinação de constituintes que consiste em apor, por intermédio de uma partícula subordinante, um constituinte da oração a outro, de tal modo que entre eles se estabeleça uma relação de regência. Quando esses liames de subordinação se posicionam *antes* da regência, isto é, *antecedem o termo regido*, temos a *preposição*; quando, ao contrário, eles se colocam *após o termo regido*, ocorre a *posposição*. Assim, o termo *preposição*, empregado para nomear uma das dez classes de palavras do português, foi cunhado por extensão com o processo de ordenação de constituintes de que participa e indica, portanto, que, nessa língua, assim como em várias outras da família indo-europeia, a regência segue imediatamente a partícula subordinante: *Cheguei ao meu destino; tenho medo de tempestade*. Essa breve reflexão acerca da etimologia do termo *preposição* já nos permite extrair, ao menos, duas generalizações importantes a respeito da classe: (i) nem todas as línguas possuem preposição; (ii) a preposição não é uma mera partícula de ligação, mas sim um operador sintático de subordinação e, como tal, uma categoria predicadora que pode, portanto, atuar como núcleo de sintagma⁴. Cabe-nos, então, analisar algumas implicações de tais generalizações. Considerando a demonstrada propriedade predicadora da preposição e sua inexistência em alguns sistemas linguísticos, há de se inquirir, primeiramente, de que modo tais sistemas expressam as relações codificadas pelas preposições em contextos semelhantes. Segundo

⁴ Eis a principal propriedade que separa a classe das preposições da classe das conjunções. “Tanto as preposições como as conjunções são palavras de ligação, são invariáveis e não podem funcionar como elementos frásicos. Contudo, as preposições distinguem-se das conjunções pela sua capacidade de regência e por serem semanticamente a expressão de uma dada relação (al. Verhältniswort: “palavra de **relação**”) (Vilela e Koch, 2001: 256, negrito dos autores, sublinhado nosso).

Saint-Dizier (2006), na ausência de preposição, as relações são codificadas por meio de marcas morfológicas, tais como morfemas de caso ou afixos, a exemplo do que ocorria no latim clássico⁵. Na avaliação desse teórico, “o fato de casos ou de outros morfemas e afixos serem empregados em algumas línguas ao invés de preposições indica que as preposições têm relações específicas com outros tipos de mecanismos linguísticos” (Saint-Dizier, 2006: 2, tradução nossa⁶). Parece ter sido essa também a intuição de Sapir, citado por Câmara Jr. (1989 [1941]: 128), para quem a preposição é um morfema que expressa uma “relação pura”, assim como “os casos nominativo e acusativo nas línguas clássicas (ditos casos “gramaticais”)”. Bagno (2011), ao tratar do tema, relaciona a riqueza dos sistemas de caso das línguas ao modo como realizam aposições. Segundo o autor,

línguas com sistemas de caso mais limitados tendem a ter somente casos gramaticais (acusativo, ergativo, genitivo, dativo, etc.) e, portanto, nenhuma aposição gramatical.⁷ [...] Assim, as línguas podem usar casos em lugar de aposições gramaticais, e nomes relacionais no lugar de aposições concretas (p. 855, itálico do autor).

Fato é que, independentemente de a língua recorrer a caso ou a nome relacional para codificar suas relações de aposição, tais relações exercem na oração funções predicativas que vão das mais concretas (Moro *em* Belo Horizonte) às mais abstratas (Estávamos *em* 1902 e não pensávamos *em* tecnologia de ponta para

⁵ Segundo Luraghi (2003:11), “cases and prepositions encode grammatical relations and semantic roles of nominal constituents. Cases are sometimes thought to be ‘more grammatical’ than prepositions [...] Especially when encoding grammatical relations, cases and prepositions are often considered purely distinctive markers, without an autonomous semantic content”. Ela observa que, mesmo no sistema de casos, existem níveis distintos de abstração: “based on the frequency with which a certain case encodes a grammatical relation or an SR [semantic relation], Indo-Europeanists traditionally distinguish between ‘grammatical’ and ‘concrete’ cases. In the classification of Kurylowicz (1964), grammatical cases are those whose primary function is to encode a grammatical relation, while concrete cases are those that most frequently encode SRs. Grammatical cases can occasionally have concrete uses for example, the Greek accusative, a grammatical case whose primary function is to encode direct object, can occasionally express Direction, i.e. encode an SR” (p. 14).

⁶ Do original: “The fact that cases or other morphemes or affixes are used in some languages instead of prepositions indicates that prepositions have specific relations with other types of linguistic mechanisms” (Saint-Dizier, 2006: p. 2).

⁷ De um ponto de vista semântico, o autor distingue *aposições gramaticais*, aquelas “que exprimem relações sintáticas muito abstratas” (p. 854), ou “puras”, nos termos de Sapir, de *aposições concretas*, “assim chamadas porque indicam sempre uma localização definida” (Bagno, 2011: 854). As aposições gramaticais são expansões das aposições concretas decorrentes de processos de gramaticalização.

tratar doenças.). Esse *continuum* de abstração se reflete conseqüentemente no tipo de relação predicativa estabelecido pelas preposições. Assim, enquanto no português há aquelas que são naturalmente dotadas da capacidade de seleção argumental, sendo, por isso, predicativas por excelência (*contra X*, por exemplo), outras há que funcionam no escopo da marcação de funções sintáticas (casos gramaticais), como o objeto indireto, por exemplo (dar *X a Y*), desempenhando funções predicativas apenas quando combinadas com predicadores verbais (ir *a X*). Decorre, portanto, dessa propriedade da classe a descrição de Dubois (1997 [1973]: 483), segundo a qual “as preposições traduzem [...] relações gramaticais e espaço-temporais”. Tal descrição, aliada à propriedade predicadora da espécie, atesta a multifuncionalidade dessa categoria linguística e evidencia que o entendimento de que as preposições são partículas de união convencionalmente esvaziadas de significação⁸ é demasiado reducionista e, como tal, equivocado. De acordo com Saint-Dizier (2006), são variados os papéis dessas “pequenas notáveis”, nos termos de Bagno (2011):

Preposições podem primeiramente ser vistas como uma **categoria funcional** na sintaxe: elas são núcleo de sintagmas preposicionados. A preposição, então, domina hierarquicamente o sintagma nominal. Preposições também podem ser vistas como uma **relação semântica** entre uma estrutura que a precede (por exemplo, um verbo) e outra que a segue (por exemplo, um sintagma nominal). Essa relação pode ser representada como uma relação conceitual [...]. Finalmente, as preposições podem ser vistas como uma **categoria lexical** que impõe tanto seleção categorial (nível da estrutura) quanto seleção semântica (restrição no nível semântico) (Saint-Dizier, 2006: 2, tradução nossa⁹, negritos do autor).

⁸ Luraghi (2003: 11) avalia que as formas gramaticais, categoria a que pertencem as preposições, também são portadoras de significado e assim as compara em relação às formas lexicais: “the substance of their meaning is not different from the substance of lexical meaning: the difference lies in the degree of abstractness, rather than in substance”, concepção de que compartilhamos. Visão semelhante tem Castilho (2010: 583), para quem “cada preposição tem um sentido de base, de localização espacial ou temporal”.

⁹ Do original: “Prepositions can first be viewed as a functional category in syntax: they are heads of prepositional phrases. The preposition then hierarchically dominates the noun phrase. Prepositions can also be viewed as a semantic relation between a structure that precedes it (e.g. a verb) and another one that follows it (e.g. an NP). This relation can be represented as a conceptual relation [...]. Finally, prepositions can be viewed as a lexical category that imposes both a categorial (structure level) and a semantic selection (semantic restriction level)” (Saint-Dizier, 2006: 2).

Castilho (2010: 583) agrega a essas funções uma quarta, ligada ao discurso: “acréscimo de informações secundárias ao texto e organização do texto, no caso das construções de tópico preposicionado”. Essa polivalência funcional decorre do fato de que a preposição transita por duas categorias – lexical e funcional –, o que lhe confere um estatuto de categoria intermediária que, se não concebida numa perspectiva multissistêmica, suscita um problema teórico, já que os atributos lexical e funcional tendem a ser evocados para opor dicotomicamente léxico e gramática, delimitando-se as preposições aos domínios da gramática em função, entre outros, do alegado e controverso esvaziamento semântico da classe. Essa visão dicotômica e polarizada do sistema linguístico não se conforma à concepção da teoria cognitivista de estudo da língua(gem). De acordo com Langacker (2009), na perspectiva da gramática cognitiva, léxico e gramática são concebidos como um *continuum* e, nessa visão, as preposições são categorias de *status* intermediário em, ao menos, três dimensões:

primeiro [...] elas são intermediárias em relação ao *continuum* léxico-gramática. Segundo, elas são intermediárias em termos de categoria gramatical, em que nome e verbo representam a oposição máxima. Finalmente, preposições têm *status* intermediário no que diz respeito à escala de distância e aos reinos ativo *versus* circunstancial¹⁰ (Langacker, 2009: 25, tradução nossa¹¹).

Além do benefício de neutralizar a aludida inconsistência teórica e de propiciar uma descrição mais condizente com a relevância do papel relacional¹² e predicativo desempenhado pela preposição, a abordagem cognitivista e multissistêmica nos permite ainda admitir que tanto o léxico quanto a gramática são passíveis de incorporações, ainda que em graus distintos de produtividade, o que

¹⁰ Na proposta do autor, pautada na experiência humana, o reino ativo é aquele da ação, da mudança e da força, onde as criaturas agem deliberadamente no mundo. Contrariamente, o reino circunstancial é aquele das configurações, das localizações e das situações estáticas, onde objetos com propriedades estáveis são organizados de modo particular.

¹¹ Do original: “First [...] they are intermediate in regard to the lexicon-grammar continuum. Second, they are intermediate in terms of grammatical category, where noun and verb represent a maximal opposition. Finally, prepositions have intermediate status with respect to the distance scale and the active vs.circumstantial realms” (Langacker, 2009: 25)

¹² Na perspectiva cognitivista, “as preposições são entendidas como expressões relacionais, já que expressam como o conceitualizador configura as partes que constituem uma cena espacial com respeito a outra” (Cifuentes Hornrubia, 2001, *apud* Castilho, 2010: 584).

implica reconhecer que as preposições não constituem uma classe aberta, tampouco estritamente fechada. Nesse sentido,

da mesma forma que as demais categorias predicativas, as preposições têm restrições de tipo em seus argumentos, atribuem papel temático e têm conteúdo semântico, possivelmente subespecificado. A única diferença em relação às outras categorias de classe aberta, como nomes, verbos ou adjetivos é que elas não têm qualquer morfologia (Saint-Dizier, 2006: 2, tradução nossa¹³),

o que lhes rende o atributo de *palavras invariáveis*. Tal invariabilidade é um resquício de seu processo de gramaticalização, já que a maioria delas são formas gramaticalizadas de antigos advérbios locativos.

1.1. Preposições essenciais e acidentais: uma categorização pautada na função

A tradicional distinção da classe entre *preposições essenciais* e *preposições acidentais*, embora não explicitamente posta, é um legado da tradição filosófica dos estudos gramaticais, mais precisamente dos conceitos de *substância* (*essência*) e de *acidente*.

Toda e qualquer língua humana, para cumprir a tarefa de simbolização e expressão do mundo e das experiências reais e/ou virtuais - e é para isso que os seres humanos desenvolveram a linguagem -, tem que poder enunciar o que é essencial e o que é acidental, o que é estático e o que é dinâmico (Bagno, 2011: 408).

Fica claro, pois, que o critério referido é de natureza semântica, pautado em traços [\pm dinâmico], [\pm estático], e não se conforma com um critério aparentemente histórico em que se apoiam alguns gramáticos que concebem como essenciais palavras que “sempre foram preposições” por oposição a palavras de outras classes gramaticais que “acidentalmente funcionam” como preposições. O critério

¹³ Do original: “Similarly to the other predicative categories, prepositions have type restrictions on their arguments, they assign thematic roles, and they have a semantic content, possibly underspecified. The only difference with the other open-class categories like nouns, verbs or adjectives is that they do not have any morphology” (Saint-Dizier, 2006: 2).

semântico seria apropriadamente aplicável, nesses casos, considerando-se que alguns integrantes da classe das preposições, em sua essência, são estáticos, no sentido de que não apresentam características funcionais que lhes permitam atuar em outra classe e, por isso, são ditos *essenciais*, ao passo que outros, em razão de seu estágio de gramaticalização, dispõem de determinadas propriedades que lhes permitem integrar ora a classe das preposições, ora a classe de onde provieram, daí serem denominados, em função dessa dinamicidade funcional, *acidentais*.

Parece-nos, contudo, ser possível refinar um pouco mais essa categorização, combinando o critério semântico ora exposto com um critério sintático, o que se mostra compatível com a multifuncionalidade da classe. Considerando se tratar, como demonstrado, de uma categoria subordinante com encargo predicativo, espera-se que a função genuína da preposição seja reger a introdução de um termo integrante, exigido que é pela predicação. Nesse caso, “o complemento natural e não marcado de uma preposição é o sintagma nominal” (Renzi *et al.*, 2001: 523, tradução nossa¹⁴), mas esse complemento pode ser satisfeito também, ainda que de forma menos produtiva, por um sintagma verbal (*gosta de estudar*), por um sintagma adverbial (*foi para dentro*) ou por um sintagma oracional (*necessidade de que o apoiem*). Além disso, dada a potencialidade da preposição para operar em diferentes ambientes sintáticos, ela pode atuar ainda fora desse escopo, introduzindo termo acessório. Essa potencialidade parece decorrer do próprio processo de combinação de constituintes de que as preposições participam, já que, de acordo com Hagège (2010), citado por Brenda (2014), por meio dele, as palavras “relacionam duas partes da sentença: o complemento de uma aposição e o *predicador* dessa sentença *ou um nome não-predicativo*” (p. 59, tradução nossa¹⁵, itálicos nossos). A duplicidade funcional decorrente de tal propriedade não é exclusiva do português, mas padrão recorrente em outras línguas românicas. Na língua italiana, os sintagmas preposicionados são classificados segundo o tipo de preposição que os encabeça. Conforme Renzi *et al.* (2001), as preposições

¹⁴ Do original: “Il complemento naturale e non marcato di una preposizione è il sintagma nominale [...]” (Renzi *et al.*, 2001: 523).

¹⁵ Do original: “Hagège (2010:1) defines adpositions as words which can form adpositional phrases and which relate two parts of a sentence: the complement of the adposition and the predicate of this sentence or a non-predicative noun” (Brenda, 2014: 59).

“próprias” ou “primárias” introduzem complemento, ao passo que as preposições ditas “impróprias”, “secundárias” ou “adverbiais” não introduzem complementos, mas usos tradicionalmente definidos como adverbiais, o que corresponderia, em termos sintáticos, aos adjuntos ou modificadores¹⁶. No português, não distinguimos as preposições em dois grupos distintos conforme sua função predicadora, como no italiano, até porque, segundo Neves (2018: 730), “todas as PREPOSIÇÕES que existem em português podem funcionar fora do sistema de transitividade, ou seja, todas elas podem introduzir adjuntos, mas nem todas funcionam no sistema de transitividade, ou seja, nem todas elas introduzem complementos¹⁷”. Diante disso, parece-nos possível dilatar o escopo da classificação *essencial* e *acidental* agregando ao traço semântico da dinamicidade o traço sintático da seleção argumental. Nessa óptica, seriam *essenciais* aquelas preposições que, para além de introduzirem adjuntos, podem introduzir complementos; por seu turno, o grupo das *acidentais* seria composto por aquelas preposições que funcionam apenas fora do sistema de transitividade, isto é, introduzindo adjuntos ou modificadores, casos em que atuariam como *transpositores*, nos termos propostos por Bechara (1999).

Num primeiro momento, essa propositura se mostra econômica do ponto de vista descritivo, por congregar uma tendência das línguas românicas na medida em que encontra eco também na sintaxe do espanhol. Nesta língua, abrigam-se sob o rótulo de *preposições imperfeitas*

um conjunto de partículas que habitualmente não se classificam dentro do grupo das preposições, mas que apresentam semelhança com esta classe

¹⁶ “Opor complementos a adjuntos é uma prática corrente nas descrições gramaticais das línguas indo-européias. A despeito das muitas dificuldades teóricas que cercam essa distinção, ela é sustentada basicamente pelo pressuposto de que, na estrutura da oração, certos constituintes - *complementos* - são necessários, isto é, ‘exigidos’ pelos termos regentes [predicadores/núcleos] (que sem eles ficariam ‘incompletos’), enquanto outros - *adjuntos* - são acessórios, isto é, acrescentados ‘facultativamente’ às construções” (Azeredo, 1999: 77). O fato de os adjuntos serem termos sintaticamente periféricos em relação aos complementos não significa que sejam supérfluos ou irrelevantes para a atividades discursiva.

¹⁷ Registre-se que a constatação de Neves (2018) dialoga com os postulados de Ilari *et al.* (2015), no âmbito do Projeto da Gramática do Português Culto Falado no Brasil, quando discutiram a capacidade da preposição para entrar em diversos ambientes sintáticos. Segundo os autores, as preposições introdutoras de argumentos são apenas aquelas que se “encontram em um grau mais avançado de gramaticalização. Essas preposições caracterizam-se ainda (i) por poderem ser amalgamadas a outros elementos linguísticos, (ii) por terem um sentido mais vago/difuso, e (iii) por conterem, comparativamente, menos material fonético [...]” (p. 178), o que também fora defendido por Bagno (2011).

gramatical, junto com algumas características que as diferenciam dela: *durante, mediante, obstante, embora, exceto, salvo, menos e incluso*. [...] Todas as partículas [...] têm em comum uma característica que as distingue do resto das preposições: não impõem o caso oblíquo aos pronomes que tomam como termo (Bosque & Demonte, 1999: 587, tradução nossa¹⁸).

Logo, essas partículas, que correspondem às nossas preposições acidentais, assim como no português, não introduzem complementos, mas tão somente adjuntos. Isso nos permite, então, generalizar que, do ponto de vista sintático, as preposições essenciais se distinguem das acidentais em virtude de estas não introduzirem complemento, o que implica assumir que as preposições essenciais podem ser predadoras, mas não as acidentais. É preciso, contudo, ponderar que o fato de as preposições acidentais só introduzirem adjuntos não implica que as preposições essenciais não o façam. Como demonstram os dados de Ilari *et al.* (2015) e de Neves (2018), no português, o traço não marcado para a sintaxe das preposições é a introdução do adjunto, enquanto o traço marcado é a introdução do complemento. Essa é uma função restrita ao grupo das preposições essenciais e, ainda assim, embora estas sejam potenciais aspirantes a ocorrerem em ambientes de complementação, nem todas dispõem das propriedades necessárias para fazê-lo. Acrescente-se a tal distinção outra dela decorrente e que incide diretamente na escolha da preposição. De acordo com Ilari *et al.* (2015: 176), “no processo de complementação, quem seleciona a preposição é a palavra ou sintagma a ser completado; no caso da adjunção, a escolha da preposição depende da natureza do adjunto”. Em termos práticos, isso significa que é muito mais fácil prever a preposição que figura em ambiente sintático de complementação, que aquela que figura no ambiente de adjunção, pois enquanto neste a “preposição é ditada pela natureza do adjunto ou pelo substantivo nele contido” (Ilari *et al.*, 2015: 175), naquele ela é determinada pelo termo regente e, portanto, sujeita às suas

¹⁸ Do original: “Bajo la denominación de ‘preposiciones imperfectas’ agrupamos aquí un conjunto de partículas que habitualmente no se clasifican dentro del grupo de las preposiciones, pero que presentan semejanzas con esta clase gramatical, junto com algunos rasgos que las diferencian de ella: *durante, mediante, obstante, embargante, excepto, salvo, menos e incluso*. [...] Todas las partículas [...] tienen em común una característica que las distingue del resto de las preposiciones: no imponen el caso oblicuo a los pronombres que toman como término” (Bosque & Demonte, 1999: 587).

restrições sintáticas e semânticas¹⁹. Essa especialização funcional e contextual, bem como a maior fixidez (+ estático) do grupo das preposições essenciais em relação às acidentais (+ dinâmico), permitem-nos acrescentar mais um traço distintivo entre os dois grupos: as preposições essenciais tendem a ser mais gramaticalizadas que as preposições acidentais, cujo processo de gramaticalização ainda está em curso, generalização que se alinha com os postulados de Bagno (2011) e de Ilari *et al.* (2015).

Discutida a proposta de subcategorização das preposições segundo a função, passaremos a discuti-la sob a perspectiva da forma. Antes, contudo, apresentaremos, para fins de sistematização do que aqui se propôs, um quadro síntese dos traços especificadores das preposições essenciais e das acidentais na língua portuguesa.

QUADRO 1 - Traços especificadores das preposições essenciais e acidentais

		PREPOSIÇÕES ESSENCIAIS	PREPOSIÇÕES ACIDENTAIS
CRITÉRIO ADOTADO	Semântico	[+ estáticas] = funcionam apenas como preposições	[+ dinâmicas] = funcionam como preposições e como outras classes gramaticais
	Sintático	[± complemento; + adjunto] = podem introduzir complemento e adjunto	[- complemento; + adjunto] = introduzem apenas adjuntos
	Continuum de gramaticalização	[+ gramaticalizadas]	[- gramaticalizadas]

(Fonte: elaborado pela autora)

¹⁹ Perini (2010) denomina de *funcionais* as preposições que ocorrem em contexto de predicação por oposição às *predicadoras*, que ocorrem em contextos de adjunção e cuja função é atribuir papel temático ao sintagma que ajudam a formar.

1.2. Preposições simples e complexas: uma categorização pautada na forma

Outro critério comumente evocado para subcategorizar as preposições é aquele baseado na forma, segundo o qual os itens da classe são classificados como unidades simples ou complexas. Nesse viés, consideram-se como preposições simples aquelas formadas de uma única palavra e, como preposições complexas – ou locuções prepositivas, nos termos tradicionais –, aquelas formadas por mais de um item lexical. Segundo Coutinho (1970 [1938]: 268), “no latim vulgar, era frequente combinar-se uma preposição com outra, resultando disso a formação de uma preposição composta”. Fato é que estabelecer o estatuto de uma locução prepositiva é tarefa bem mais complexa que essa descrição diacrônica possa fazer crer. De acordo com Bosque & Demonte (1999: 579), a locução prepositiva, em um sentido amplo, pode ser definida

como uma expressão constituída por várias palavras, com uma forma fixa, que se utiliza na fala como uma peça única e que apresenta o comportamento típico de uma preposição (expressa o mesmo tipo de relações, aparece nos mesmos contextos sintáticos e alterna com preposições de significado similar) (tradução nossa²⁰).

Em sentido restrito, a locução deve constituir uma verdadeira unidade léxica pertencente à categoria das preposições. “Nesse caso, não possuirá uma estrutura interna produtiva, para a qual duas exigências devem ser satisfeitas: a de fixação ou invariabilidade e a de coesão ou inseparabilidade” (Bosque & Demonte, 1999: 579, tradução nossa²¹). Fica evidente, pois, que a definição de sentido amplo busca articular o conceito tradicional de *locução* com critérios de natureza semântica, sintática e prosódica que envolvem a classe das preposições, enquanto a de sentido restrito tem sua essência na coesão da forma. O problema que se coloca é que, se

²⁰ Do original: “Podemos definir el concepto de ‘locución prepositiva’ como una expresión constituída por varias palabras, con una forma fija, que se utiliza em el habla como una pieza única y que presenta el comportamiento típico de una preposición (expresa el mismo tipo de relaciones, aparece em los mismos contextos sintáticos y alterna com preposiciones de significado similar” (Bosque & Demonte, 1999:579).

²¹ Do original: “En este caso, no poseerá una estructura interna produtiva, para lo cual deben cumplirse dos requisitos: el de la fijación o invariabilidad y el de la cohesión o inseparabilidad” (Bosque & Demonte, 1999: 579).

tais formulações são suficientemente amplas para estabelecer com segurança quais construções podem ser categorizadas como locuções prepositivas no espanhol, o mesmo não se pode admitir em relação ao português. No tocante ao critério sintático, como demonstrado na subseção precedente, mesmo em se tratando de preposições simples, existem algumas restrições, já que, no português, as preposições essenciais atuam em ambiente de transitividade, ao contrário das acidentais. Logo, tal critério seria válido apenas por aproximação, limitando-se à função de elo relacional, o que neutraliza, evidentemente, a distinção com a classe das conjunções. Além disso, ao que parece, as preposições complexas do português, assim como as preposições acidentais, não podem figurar em contextos de complementação, a não ser naqueles de complementos circunstanciais:

- (1) a. Minha mãe mora em frente à igreja.
b. * Minha mãe acredita em frente à igreja.

Essa restrição contextual pode ser explicada pela identidade funcional que aproxima o complemento circunstancial do grupo dos advérbios e das locuções adverbiais²², que, no português, são a base prototípica das locuções prepositivas, tal como defendido por Said Ali (1965) e por Bechara (1999: 301):

Locução prepositiva é o grupo de palavras com valor e emprego de preposição. Em geral a locução prepositiva é constituída de advérbio ou locução adverbial seguida da preposição *de*, *a* ou *com*. [...] Às vezes, a locução prepositiva se forma de duas preposições, como: *de per* (na locução *de per si*), *até a* e *para com*. [...]

OBSERVAÇÃO: O substantivo que às vezes entra para formar estas locuções em geral está no singular; mas plural também é possível: *Viver à custa do pai* (ou *às custas do pai*), *O negócio está em via de solução* (ou *em vias de solução*).

A definição apresentada por Bechara (1999) dialoga intimamente com aquela de sentido amplo proposta por Bosque & Demonte (1999) e, portanto, esbarra na já discutida limitação do critério sintático. No que tange ao critério

²² Ressalte-se que raciocínio análogo levou a Nomenclatura Gramatical Brasileira a abolir o complemento circunstancial, à revelia da natureza predicadora do verbo que o requer, incluindo-o no bojo dos adjuntos adverbiais, conduta endossada por muitos gramáticos normativos.

semântico, a limitação de ambas se ancora na dificuldade de se precisar um sentido que seja comum a todas as preposições, o que poderia ser relativizado, contudo, considerando-se a similaridade de significação referida. Outra solução paliativa, nesse sentido, seria, tal como proposto por Ilari *et al.* (2015: 292),

estender às locuções prepositivas o sentido mais geral a ser atribuído à classe das preposições [...] o de indicar relações espaciais. Se assim procedermos, teremos que classificar como locuções prepositivas as locuções que designam relações espaciais ou relações metafóricas derivadas de relações espaciais.

A despeito de, tal como defendem os autores, essa extensão de valor semântico possibilitar uma descrição bastante rica das relações espaciais do sistema linguístico, não podemos desconsiderar sua insuficiência para permitir estabelecer um estatuto categorial próprio das locuções prepositivas, na medida em que as aproxima, por exemplo, de locuções adverbiais que também perfilam noções espaciais. Uma possível forma de refiná-la seria, talvez, combiná-la com a distribuição de seus elementos constituintes, já que as locuções adverbiais não requerem, como as prepositivas, uma preposição como último elemento do conjunto, conforme ilustram, respectivamente, os exemplos (2) e (3):

(2) O local do acidente fica logo à frente.

(3) Frente ao impasse, a votação da reforma da previdência foi postergada.

Embora, à luz dos exemplos arrolados, o problema possa parecer de fácil solução, o expediente de combinar critério semântico com critério formal de distribuição mostra-se parcialmente eficaz para assegurar a categorização. Segundo adverte Pontes (1992: 50),

o fenômeno é mais complexo do que simplesmente a locução terminar ou não com Prep [preposição]. Existe a possibilidade de elipse do SN, quando o contexto do discurso (seja dentro da S [sentença], seja em ambientes maiores do que S) ou da situação (anáfora ou pragmática) supre a informação. Quando se dá a elipse do SN, a Prep. é elidida

também, porque em português não é geralmente permitido sobrar a Prep. (a não ser em casos especiais), como pode acontecer em inglês.

Outro complicador que se observa é quando, no constituinte em análise, existe um nome que indica lugar, tal como ilustrado nos exemplos a seguir, extraídos de Pontes (1992: 47), com sublinhados nossos:

- (4) “Maria está do lado de Joana”.
- (5) “O livro está do lado esquerdo da casa”.
- (6) “Maria foi lá para os lados de Jacarepaguá”.
- (7) “Coloquei o livro de lado²³”.

Nesses casos, mesmo com a presença da preposição, não é possível reconhecer a locução prepositiva em nenhum dos contextos, fato que leva Pontes (1992: 47) à seguinte formulação:

constatamos que, nas listas de L. Preps. [locuções prepositivas] dos gramáticos, temos graus diferentes de evolução de N, que trazem como consequência o problema de se decidir se se tem uma S. Prep. [sequência preposicionada] formada de Prep. + SN, ou se se tem uma L. Prep. Penso que, conforme o grau de evolução, não se deve considerar a seqüência como L. Prep.

O grau de evolução a que se refere Pontes (1992) corresponde ao estágio de gramaticalização das formas, o que, incontestavelmente, irá interferir tanto na invariabilidade quanto na inseparabilidade da sequência. Segundo observou essa linguista, no português, “há diferentes graus de coesão nas sequências, conforme as palavras sejam mais ou menos invariáveis. Nos casos em que a evolução do N já chegou ao máximo, a coesão da locução foi tal que os gramáticos escrevem como

²³ A autora reconhece que “de lado”, embora tenha uma conotação espacial, difere das demais combinações por indicar, no contexto, mais um modo de colocar o livro que o lugar em que é colocado. Isso mostra que os nomes locativos apresentam graus distintos de conotar a locação, o que, num *continuum*, os aproxima ora da categoria *nome*, ora da categoria *advérbio*. Sem dúvida, essa gradação está diretamente associada ao processo metafórico da gramaticalização, que também envolve ambiguidade. A sentença em (9), sobretudo extraída de seu contexto maior, apresenta uma leitura ambígua, dado que permite ainda compreender a expressão “de lado” não como um lugar físico, mas como um lugar mental, significando à parte, interpretação que se assemelha com a proposta de Talmy (2005) segundo a qual a categoria cognitiva *espaço* pode compreender tanto um espaço físico quanto um espaço mental.

uma palavra só²⁴ (p. 45). Dado que, conforme enunciado no conceito restrito de locução prepositiva formulado por Bosque & Demonte (1999), no espanhol, a coesão é condição imperativa para se estabelecer uma locução prepositiva, podemos inferir que as locuções prepositivas nesse idioma tendem a ser mais gramaticalizadas que aquelas existentes no português, o que nos faz pressupor que, por conseguinte, elas formam uma classe mais aberta na língua de Camões que na de Cervantes.

Para além das particularidades já destacadas, a distribuição dos elementos constituintes das locuções prepositivas parece ser outro fator que diferencia/aproxima as línguas românicas. Enquanto no italiano e no espanhol essas locuções formam-se prioritariamente com uma base nominal, no português, como demonstrado, a base prototípica é adverbial. No italiano, "a forma canônica de locução preposicional é 'preposição + nome + SP': *a confronto di, in confronto a, di fronte a, per mezzo di, in mezzo a, in base a, a paragone com, per opera di, per via di, a causa di, a danno di, etc.*" (Renzi et al., 2001: 544, tradução nossa²⁵). No espanhol, as locuções prepositivas de base nominal "respondem a dois tipos de estrutura: <N + P> (*frente a; cara a; etc.*) e <P + N + P> (*com relación a; a propósito de; etc.*)" (Bosque & Demonte, 1999: 579, tradução nossa²⁶). Registram-se ainda, em menor grau de produtividade, aquelas em cuja base se encontram elementos distintos de nome substantivo: (i) adjetivo/particípio + preposição (*conforme a, debido a, etc.*); (ii) advérbios combinados com preposição (*fuera de, por entremedias de, etc.*); (iii) combinações de preposições (*por bajo de; por contra de*); (iv) combinações de verbos e preposições (*a juzgar por, a partir de*); (v) como seguido de substantivo (*como consecuencia de, como resultado de*). A despeito de as locuções do italiano e do espanhol se construïrem em torno de uma base

²⁴ Ilari et al. (2015) não relacionam a forma de escrita à coesão da forma, tal como Pontes (1992). Segundo esses autores, "algumas locuções já se haviam consolidado quando a gramática e a ortografia fizeram suas sistematizações da língua, e é por isso que escrevemos como uma só palavra *perante* e *desde*, mas separamos "*para com*", ou que escrevemos em duas palavras "*apesar de*" e em três "*a despeito de*" (p. 299), visão que também endossamos.

²⁵ Do original: "La forma canônica di locuzione preposizionale è 'preposizione + nome + SP': *a confronto di, in confronto a, di fronte a, per mezzo di, in mezzo a, in base a, a paragone com, per opera di, per via di, a causa di, a danno di, etc.*" (Renzi et al., 2001: 544).

²⁶ Do original: "[Dichas locuciones] responden a dos tipos de estructura: <N + P> (*frente a; cara a; etc.*) e <P + N + P> (*com relación a; a propósito de; etc.*)" (Bosque & Demonte, 1999: 579).

nominal, o espanhol se distingue do italiano por admitir locução não introduzida por preposição e se aproxima do português naquelas locuções cuja base não é de natureza substantiva. Entretanto, no português, além da distinção quanto à prototipicidade da base, as regras de combinação são mais restritivas, sobretudo naquelas construções em que o vocábulo que combina com a preposição fica invariável e que, portanto, são mais gramaticalizadas:

há uma divisão nítida entre as palavras que ocorrem nas diferentes locuções: há uma classe de palavras que podem ser precedidas da Prep. *em*, e outras que não admitem ser antecedidas desta Prep. Há uma coincidência, que é o fato de as palavras que não podem ser antecedidas de *em* serem invariáveis e poderem ser classificadas como advérbio de lugar. Já as que admitem *em* são Ns: ou substantivos (como *lado*, *frente*, *redor*) ou adjetivo como *baixo*. [...] Fica claro, portanto, que há uma diferença de classes, aí. Há locuções formadas de *em* + N e locuções formadas das outras Preps. + Adv. Em outros casos, a Prep. *de*, final, é opcional (Pontes, 1992: 55).

Fica evidente, a partir da descrição de Pontes (1992), que esse grupo de locução prepositiva do português altamente gramaticalizado é sempre introduzido por preposição e que o subgrupo que tem por base um constituinte nominal, estrutura canônica no espanhol e no italiano, requer a preposição *em*, ao passo que as demais preposições introduzem o subgrupo constituído de base adverbial locativa, motivo pelo qual essa sequência é mais produtiva e, portanto, a canônica na língua portuguesa. Além disso, mesmo no grupo das locuções mais gramaticalizadas, existe a opcionalidade de emprego da preposição final *de*, o que coloca em xeque, como demonstrado, a adoção desse critério para distinguir entre locução prepositiva e adverbial.

2. Proposta de categorização das preposições segundo um critério cognitivo e multissistêmico

Para levarmos ao termo a tarefa a que nos propusemos nesta seção, estamos assumindo que “as palavras, assim como as funções sintáticas, compõem um *quase-continuum*, em oposição à classificação discreta tradicional” (Mello, 1990: 128). Essa

assunção implica admitir que, nesse *continuum*, existem itens que congregam um maior conjunto de traços definidores da categoria que outros e, conseqüentemente, que tais elementos são, por isso, não só mais marcados, como também mais representativos da classe, o que se alinha à noção de protótipo da teoria cognitiva. Ocorre que, para determinarmos qual exemplar é mais nuclear e qual é mais periférico na classe das preposições, necessitamos, primeiramente, definir uma matriz de traços que nos permita comparar o conjunto de itens em termos desse feixe de traços e fixar, com base em critérios quantitativos de preenchimento da matriz, quais são mais representativos da categoria e quais são mais periféricos, assim considerados por congregarem o menor número de traços. No intuito de obter maior alcance descritivo, nossa matriz será construída com base em oito critérios que se distribuirão, inicialmente, por três macro-critérios: (i) categorial, (ii) sintático, (ii) cognitivo. O macro-critério categorial será definido com base no maior grau de gramaticalização da forma, para cujo estabelecimento nos apoiaremos em Bagno (2011) e em Ilari *et al.* (2015). O critério sintático levará em consideração a natureza transitiva do item, isto é, sua potencialidade para atuar como complemento ou como adjunto. O macro-critério cognitivo envolve tanto representações cognitivas ligadas à forma básica das preposições em seu percurso diacrônico, para cuja definição nos apoiaremos em Said Ali (1966 [1921]), em Coutinho (1970 [1938]) e em Cunha (1997 [1982]), quanto aquelas ligadas à transposição de esquemas imagéticos de um domínio a outro, ou seja, o uso metafórico das preposições, para o que seremos subsidiados pelo estudo de Ilari *et al.* (2015). Elucidados os critérios adotados, apresentamos, a seguir, a matriz proposta para estabelecermos os exemplares prototípicos da classe das preposições. Esclarecemos que os exemplares dispostos em nossa matriz se restringem ao conjunto das onze preposições essenciais que são ainda de uso frequente na língua portuguesa contemporânea²⁷.

²⁷ Das dezesseis preposições essenciais ordenadas alfabeticamente em nossos compêndios gramaticais - “a, ante, até, com, contra, de, desde, em, entre, para, perante, por, sem, sob, sobre, trás” (Bechara, 1999: 301) - apenas onze são ainda de uso corrente na contemporaneidade e, segundo dados de Ilari *et al.* (2015), somente quatro - *de, em, para* e *a* - respondem por aproximadamente 80% do total de ocorrências da classe no *corpus* do Projeto da Norma Urbana Culta (NURC). Nossa matriz contempla, assim, apenas as onze preposições que obtiveram frequência superior a 1% no estudo de Ilari *et al.* (2015), que analisou 5.215 ocorrências de preposições extraídas do Projeto NURC. A preposição *sobre*, que pode nos parecer um exemplar prototípico da classe, não alcançou

QUADRO 2 - Matriz de traços definidores da classe das preposições

	CRITÉRIO CATEGORIAL	CRITÉRIO SINTÁTICO		CRITÉRIO COGNITIVO				
		Potencial de transitividade		Representação espacial básica		Uso metafórico		
		+ complemento	+ adjunto	+ posição no espaço	+ deslocamento no espaço	uso espacial	uso temporal	outros usos
A	x	x	x	x	x	x	x	x
ATÉ			x		x	x	x	x
COM	x	x	x	x			x	x
CONTRA			x		x			x
DE	x	x	x	x	x	x	x	x
DESDE			x		x	x	x	x
EM	x	x	x	x	x	x	x	x
ENTRE			x	x		x	x	x
POR	x	x	x		x	x	x	x
PARA	x	x	x		x	x	x	x
SEM			x	x				x

(Fonte: elaborado pela autora)

A análise da matriz acima fornece-nos um mapeamento amplo da classe das preposições e evoca uma série de reflexões, quer voltemos o olhar para as marcações dispostas horizontalmente, quer para aquelas arranjadas verticalmente. Iniciando nossa análise pela disposição horizontal de atendimento aos critérios, percebemos que três preposições (A- DE - EM) preenchem integralmente a matriz, o que significa que são exemplares mais marcados, no sentido de que congregam todos os traços relacionados como definidores da classe. Podem, por isso, ser considerados os itens mais prototípicos da categoria. Curiosamente, esses exemplares estão entre os quatro mais frequentes no *corpus* do Projeto NURC, segundo cômputo de Ilari *et al.* (2015)²⁸.

esse percentual, o que nos levou, por fidelidade ao critério metodológico adotado, a excluí-la da análise na matriz proposta. A produtividade desses onze exemplares sugere que está havendo uma especialização de uso dessas preposições na língua, o que indicia uma acentuação de seu grau de gramaticalização ou, quem sabe, uma proximidade de traços capazes de torná-las exemplares mais prototípicos da categoria.

²⁸ Essa coincidência é, de certo modo, previsível, considerando-se que existe uma correlação estrita entre frequência do item e prototipia.

No *continuum* de prototipia que buscamos estabelecer, dois outros exemplares aproximam-se bastante dos três mais prototípicos, diferenciando-se deles apenas por não marcarem, na sua forma etimológica mais básica, uma posição no espaço, mas um deslocamento ou movimento no espaço²⁹. Trata-se das preposições PARA, que completa o quarteto mais frequente no *corpus* de Ilari *et al.* (2015), e POR. Tomando por parâmetro tanto a proposta de Castilho (2010) para definir a categoria, quanto sua frequência superior à da preposição A nos dados de Ilari *et al.* (2015)³⁰, a preposição PARA poderia ser incluída entre o subconjunto dos itens mais prototípicos da classe. Nossa opção, contudo, até mesmo para nos mantermos fiéis aos critérios propostos em nossa matriz, é considerá-la, na perspectiva do *continuum* que estamos adotando, como um item que gravita em torno do núcleo numa relação de intersecção bastante estreita com ele, o que lhe permite, inclusive, a concorrência com as formas mais nucleares EM e A, mas que apresenta um matiz que o afasta dos demais elementos em alguma medida na linha do *continuum*. É possível que esse matiz da não marcação de uma posição no espaço decorra do fato de que PARA é uma pseudo-forma simples, dado que seu processo de formação decorre da junção de duas outras preposições (*per + ad*)³¹, mas em virtude também dessa especificidade acreditamos que PARA destoa minimamente do grupo das três formas eleitas como protótipos da categoria. Nesse viés, POR alinha-se ao grupo, já que é, assim como seus demais elementos, uma forma simples desde a origem, mas também destoa dele tanto pela marcação negativa do traço *posição no espaço* em sua forma etimológica mais básica, quanto

²⁹ Embora tanto a posição quanto o deslocamento no espaço denotem noções ligadas ao domínio básico da experiência humana, o deslocamento no espaço é uma noção mais abstrata que a posição no espaço, já que envolve movimento de um ponto A para um ponto B, experiência corpórea que motiva operações cognitivas de natureza metafórica. Assim, um exemplar que conota tanto noções concretas quanto mais abstratas do domínio espacial tendem a ser mais prototípicos que aqueles que conotam apenas noções abstratas, porque congregam mais traços da classe.

³⁰ Na pesquisa quantitativa empreendida por Ilari *et al.* (2015: 164) no *corpus* compartilhado do NURC, “as preposições têm uma frequência de uso muito desigual: das 5215 ocorrências registradas [...], 31% pertencem a *de*, 27% a *em*, 12% a *para* e 10% a *a*. [...] As próximas preposições, em ordem de frequência, são *com* (com 8%), *por* (com 5,5%) e *até* (com 1,7%)”.

³¹ O fato de a preposição PARA constituir uma pseudo-forma simples muito provavelmente não faz parte do conhecimento linguístico do falante contemporâneo e não foi tomado na matriz como critério para excluí-la do núcleo básico mais prototípico da classe. Tal fato histórico foi apenas evocado como uma explicação possível para essa preposição conotar apenas uma noção espacial mais abstrata, que é o deslocamento no espaço, prescindindo de marcar, como as demais do núcleo básico, a posição no espaço.

pela baixa frequência de uso, segundo dados de Ilari *et al.* (2015). Entendemos, assim, que deva figurar, no *continuum*, ao lado da preposição PARA, com a qual concorre em alguns contextos. Tomando a concorrência das formas em contextos sintáticos como mais um parâmetro de análise, há que se considerar que a preposição POR não concorre com as formas A, EM e DE, consideradas prototípicas, o que constitui mais um argumento em favor do fato de que não existe, entre elas, equivalência de traços capazes de promover a migração de POR para o conjunto dos exemplares mais prototípicos.

O próximo item da classe que reuniu o maior número de critérios, marcando positivamente seis dos oito propostos, é a preposição COM, que, dada sua origem indo-europeia e, portanto, seu longo processo de mudança linguística, hoje é bastante opaca na conotação de noções espaciais, tendo se especializado na denotação de noções relacionais. Pontuando cinco dos critérios propostos, o que, na perspectiva do *continuum*, coloca-as como itens mediais da categoria, aparecem as preposições ATÉ, DESDE e ENTRE, distintas entre si, em termos etimológicos, na especificidade para assinalar movimento (*até* e *desde*) e posição no espaço (*entre*). As duas preposições que traduzem movimento o fazem de perspectivas distintas, já que, no caso de DESDE, o foco é no ponto de partida e, no caso de ATÉ, o alvo é o ponto de chegada, o que permite seu uso em correlação na língua sem incorrer em contradição:

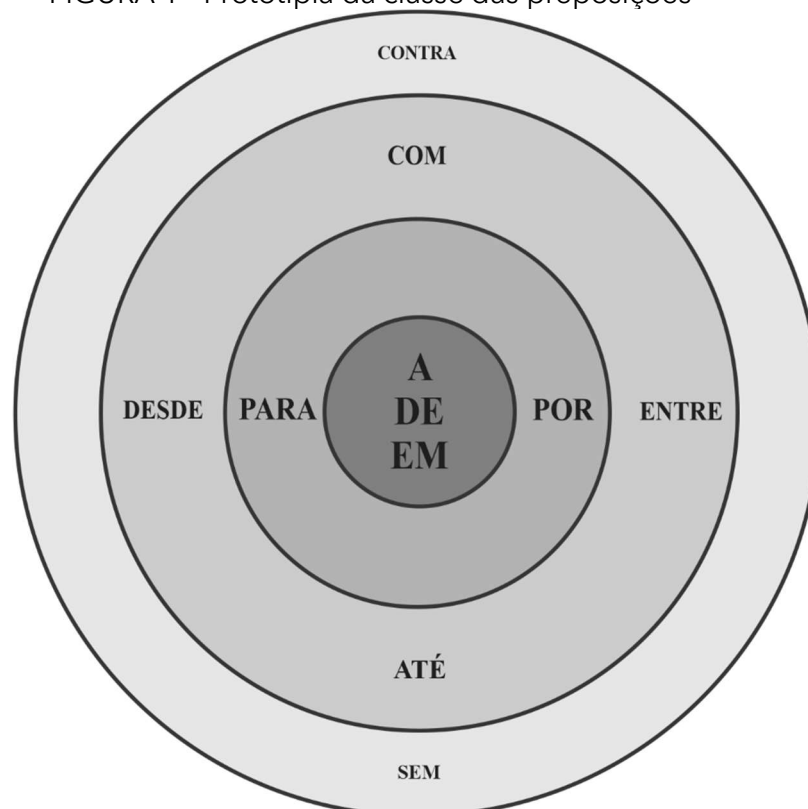
- (8) *Desde* ontem até a data do julgamento os manifestantes permanecerão em vigília.

Os exemplares mais marginais da categoria, computando apenas três e dois critérios, são, respectivamente, as preposições SEM e CONTRA. Aquela, em função de já no latim não se prestar à conotação espacial, especializou bastante seu contexto de uso, restringindo-se, quase que exclusivamente, a marcar privação ou carência. A preposição CONTRA também está bastante restrita aos contextos de oposição e, nos dados de Ilari *et al.* (2015: 197), foi encontrada ainda, apesar de pouco frequente, “em contextos em que marca um *alvo* que não é nem espacial nem temporal, mas em algum sentido psicológico: Eu não tenho nada assim

peçoal *contra* a televisão”. Outro aspecto digno de menção é que, apesar de ser considerada pouco gramaticalizada tanto por Bagno (2011) quanto por Ilari *et al.* (2015), essa partícula é bastante produtiva como prefixo na língua: *contradizer*, *contra-argumentar*, *contra-ataque*, *contrassenso...*, o que pode indiciar o término de um ciclo de gramaticalização, na medida em que está se tornando um afixo. Semelhante função, ainda que com menor produtividade, pode ser identificada também no caso de SEM: *sem-vergonha*, *sem-teto*, *sem-número*, *sem noção...* Avaliamos que essa similaridade funcional seja mais uma evidência em prol do argumento de que são exemplares mais marginais da classe das preposições, conforme estamos propondo.

No intuito de tornar mais visual nossa proposta de categorização da classe das preposições, apresentamos, a seguir, uma figura, na qual procuramos representar graficamente a prototipia proposta:

FIGURA 1 - Prototipia da classe das preposições



(Fonte: elaborado pela autora)

Definido o *continuum* de protótipos da classe das preposições básicas, função precípua da matriz analisada e cujos resultados culminaram na imagem

acima, cabe também um exame dessa classe à luz de uma verticalização dos critérios propostos na referida matriz. Iniciando pelo critério semântico, atestamos o sentido básico espacial da categoria e avaliamos ser possível tornar mais econômica a formulação de Castilho (2010), restringindo-a à proposição de que o sentido de base das preposições, pelo menos dos exemplares mais prototípicos da classe, é reconhecível quando expressam as categorias posição ou movimento no espaço. Os dados dispostos na matriz evidenciam, ainda, que as preposições que cumulam a função de traduzir tanto a posição quanto o movimento no espaço são os exemplares mais prototípicos, o que sugere que a conotação da posição no espaço (função + concreta) é, na perspectiva do *continuum*, o traço mais básico e, portanto, o mais representativo da categoria. Assim, aqueles exemplares que, no seu curso de gramaticalização, preservaram esse traço mostram-se sincronicamente mais salientes.

No tocante ao critério cognitivo, observamos que o uso espacial de base da classe, apesar de bastante produtivo, não pode ser identificado nas preposições *contra* e *sem*, exatamente as mesmas que, conforme demonstramos, estão se tornando uma categoria morfológica produtiva na função de afixo. Em contrapartida, usos outros que não os espaciais e temporais são os mais produtivos, alavancados, segundo entendimento de Bagno (2011: 862), pelas preposições mais gramaticalizadas, que “têm um espectro semântico muito mais amplo e difuso que as outras, o que lhes permite funcionar de maneira mais versátil como elementos juntores, adjuntos e complementadores”. A formulação de Bagno (*op. cit.*) contraria, contudo, os pressupostos da gramaticalização, mais especificamente no que propõem os princípios da especialização e da persistência formulados por Hopper (1991) e que são endossados pelos dados de nossa matriz. Combinando o critério categorial com o critério sintático, verificamos uma correlação biunívoca entre o maior grau de gramaticalização da preposição e seu potencial transitivo, já que apenas as preposições mais gramaticalizadas podem introduzir complementos. Isso comprova uma especialização de função sintática no âmbito da classe das preposições essenciais, já que nem todas estão aptas a figurar em ambientes de complementação, embora todas possam ocorrer em ambiente de adjunção.

Uma análise mais incisiva do critério sintático faz emergir, ainda, outras questões relevantes, na medida em que confirma a tese de que todas as preposições essenciais podem introduzir adjunto, mas restringe a introdução do complemento às preposições mais gramaticalizadas, limitando, em alguma medida, a proposta formulada no quadro 01 para distinguir preposições essenciais de preposições acidentais, sobretudo se interpretada numa perspectiva meramente dicotômica. Segundo tal proposta, as preposições essenciais são mais gramaticalizadas que as acidentais e, ao contrário destas, são potencialmente introdutoras de complementos. O refinamento da análise propiciado pela matriz de traços evidencia, pois, que a perspectiva polarizada de definição dos traços precisa ser lapidada, de modo a contemplar o intervalo compreendido entre os dois pólos. Desse modo, considerando-se a perspectiva do *continuum* com a qual estamos trabalhando, temos que admitir que, mesmo no rol das preposições essenciais, cujo grau de gramaticalização é efetivamente maior que o das acidentais, existe um grupo que está mais gramaticalizado, o que lhe faculta figurar em ambiente sintático de complementação, funcionando, assim, como um morfema que traduz uma relação sintática abstrata ligada à marcação de casos gramaticais. Esse dado atesta, portanto, uma correlação entre maior grau de gramaticalização e ambiente de complementação e, conseqüentemente, uma especialização sintática no rol das preposições essenciais. Os dados dispostos na matriz, cotejados com aqueles da figura 01, nos permitem perceber que apenas as preposições prototípicas e aquelas que ocupam a zona intermediária do *continuum* de prototipia da classe, quais sejam, A, DE, EM, PARA, POR e COM, podem figurar em contextos de complementação, o que confirma a existência de um potencial funcional distinto entre as duas subcategorias de preposição, uma vez que as preposições acidentais não apresentam potencial para figurar em contextos de complementação, restringindo-se àqueles de adjunção.

Considerações finais

As reflexões aqui empreendidas acerca das preposições demonstram algumas idiosincrasias dessa classe que, ao contrário das demais formas

linguísticas, constitui uma categoria intermediária no *continuum* léxico > gramática. Semelhante estatuto atesta não apenas seu potencial predicativo e relacional, como também a necessidade de se concebê-la numa perspectiva multissistêmica, evitando, assim, as subcategorizações dicotômicas. De tudo que aqui se discutiu, fica claro que está havendo uma especialização de usos na classe dos vocábulos tradicionalmente concebidos como preposições básicas ou essenciais, o que contribui para a redução dos itens da categoria. A análise empreendida acusou que, além da concorrência entre certas formas, que culmina inevitavelmente na eleição de algumas em detrimento de outras, tal especialização se dá tanto no plano semântico, quanto no plano sintático e no morfológico. Assim, enquanto alguns exemplares da categoria marcam tanto posição quanto movimento no espaço, outros denotam apenas uma dessas funções; além disso, apenas um conjunto restrito desses exemplares figura em contexto de complementação, o que se alinha ao maior grau de gramaticalização do item; há, também, aqueles que já se encaminham para um estágio mais avançado do processo de gramaticalização, funcionando, assim, como uma categoria morfológica que faculta ao falante ampliar o léxico por meio da formação de novas palavras, qual seja, o afixo.

Isso mostra que, de fato, a classe das preposições é semiaberta, no sentido de que, tal como observam Ilari *et al.* (2015: 299), diante da dificuldade de se criarem novas preposições, “as necessidades expressivas que poderiam ser cobertas por novas preposições vêm sendo supridas há muito tempo pela criação de locuções em que algum substantivo tem um papel essencial”. Essa dinâmica de renovação expressiva da língua pautada numa base nominal, se comparada com o padrão observado no italiano e no espanhol, leva à conclusão de que esse é um expediente fértil que aproxima as línguas do tronco românico, ao mesmo tempo em que sinaliza que a criação de locuções prepositivas a partir de bases adverbiais é um recurso linguístico já não tão produtivo no português, como registrado outrora, quando essa produtividade foi responsável por definir seu padrão canônico atestado por Said Ali (1965) e por Bechara (1999). Esse movimento diacrônico da língua, contudo, é não só legítimo, como também expectado, dado tratar-se de um sistema vivo e complexo. Nessa dinâmica, ao lado das locuções prepositivas, vocábulos integrantes de outras categorias também têm funcionado, às vezes,

como preposições, o que sinaliza que a subcategoria das preposições acidentais é o *locus* da expansão da categoria, ao passo que a subcategoria das preposições essenciais parece, de fato, fechada.

As generalizações aqui alcançadas evocam, indubitavelmente, outras questões de pesquisa, entre as quais, por exemplo, (i) se as preposições acidentais se conformam, em alguma medida, à matriz de traços proposta para as preposições básicas e, em caso afirmativo, se constituem exemplares mais ou menos periféricos; (ii) se haveria alterações no *continuum* de prototipicidade das preposições básicas quando estas integram locuções prepositivas ou quando se combinam com verbos com múltiplas valências. Isso demonstra o potencial investigativo que as preposições representam e nos convida a aprofundar o estudo dessas “pequenas notáveis”.

Referências

- Azeredo, J. C. de. (1999). *Iniciação à sintaxe do português* (5ª ed.). Jorge Zahar Editor.
- Bagno, M. (2011). *Gramática pedagógica do português*. Parábola.
- Bechara, E. (1999). *Moderna Gramática Portuguesa* (37ª ed.). Lucerna.
- Bosque, I., & Demonte, V. (1999). *Gramática descriptiva de la Lengua Española: sintaxis básica de las clases de palabras*. Espasa Calpe.
- Brenda, M. (2014). *The cognitive perspective on the polysemy of the English spacial preposition OVER*. Cambridge Scholars Publishing: Unabridged edition.
- Câmara Jr., J. M. (1989). *Princípios de linguística geral, como introdução aos estudos superiores da língua portuguesa* (7ª ed.). Padrão Livraria Editora.
- Castilho, A. T. de. (2007). Abordagem da língua como um sistema complexo. Contribuições para uma nova Linguística Histórica. In A. T. de Castilho, M. A. Torres Morais, R. E. V. Lopes, & S. M. L. Cyrino (Orgs.), *Descrição, história e aquisição do português brasileiro. Homenagem a Mary A. Kato* (pp. 329-360). Pontes / Fapesp.
- Castilho, A. T. de. (Org.). (2009). *História do português paulista*. Série Estudos (vol. 1). Instituto de Estudos da Linguagem, Unicamp.
- Castilho, A. T. de. (2009a). An approach to language as a complex system. New issues in Historical Linguistics. In A. Vanderi (Org.), *Para a História do Português Brasileiro* (Vol. VII: Vozes, veredas, voragens) (pp. 1-44). Editora da Universidade Estadual de Londrina (Republicado em A.T. de Castilho - Org. 2009, 119-136).
- Castilho, A. T. de. (2009b). Para uma análise multissistêmica das preposições. In A. T. de Castilho, *História do português paulista*. Série Estudos (vol. 1) (pp. 279-332). Instituto de Estudos da Linguagem, Unicamp.
- Castilho, A. T. de. (2010). *Nova Gramática do Português Brasileiro*. Editora Contexto.

- Coutinho, I. de L. (1970). *Pontos de gramática histórica* (6ª ed.). Livraria Acadêmica.
- Cunha, A. G. da. (1997). *Dicionário Etimológico Nova Fronteira da Língua Portuguesa*. Nova Fronteira.
- Dubois, J. (1997). *Dicionário de linguística*. Cultrix.
- Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge University Press.
- Ilary, R., Castilho, A. T. de, Leitão, M. L., Kleppa, L., & Basso, R. M. (2015). A preposição. In R. Ilary (Org.), *Gramática do português culto falado no Brasil. Volume IV: palavras de classe fechada* (pp. 163-310). Contexto.
- Hopper, P. (1991). On some principles of grammaticalization. In E. C. Traugott, & B. Heine (Orgs.). *Approaches to grammaticalization* (pp. 17-36). John Benjamins.
- Jackendoff, R. (1983). *Semantics and Cognition*. MIT Press.
- Jackendoff, R. (1987). *Consciousness and the Computational Mind*. MIT Press.
- Jackendoff, R. (1990). *Semantic Structures*. MIT Press.
- Jackendoff, R. (1992). *Languages of the Mind*. MIT Press.
- Langacker, R. (1987). *Foundations of Cognitive Grammar* (vol. I). Stanford University Press.
- Langacker, R. (1991). *Foundations of Cognitive Grammar* (vol. II). Stanford University Press.
- Langacker, R. (2009). Refleções on the functional characterization of spatial prepositions. *Belgrade English Language and Literature Studies (Belgrade BELLS)*, 1, 9-34.
- Luraghi, S. (2003). *On the meaning of prepositions and cases: the expression of the semantic roles in Ancient Greek*. John Benjamins Publishing Company.
- Martelotta, M. E., & Palomanes, R. (2009). Linguística cognitiva. In M. E. Martelotta (Org.), *Manual de linguística* (pp. 177-192). Contexto.
- Mello, H. (1990). *Adjjetivos - categoria fronteira* [Dissertação de Mestrado não publicada]. Universidade Federal de Minas Gerais.
- Neves, M. H. de M. (2018). *A gramática do português revelada em textos*. Editora UNESP.
- Oliveira, R. P. de. (2001). Semântica. In F. Mussalim, & A. C. Bentes (Orgs.). *Introdução à linguística: domínios e fronteiras* (vol. 2). Cortez.
- Perini, M. A. (2010). *Gramática do Português Brasileiro*. Parábola Editorial.
- Pontes, E. (1992). *Espaço e tempo na Língua Portuguesa*. Pontes.
- Renzi, L., Salvi, G., & Cardinaletti, A. (2001). *Grande gramática italiana di consultazione* (vol. 1). Società editrice il Mulino.
- Said Ali, M. (1966). *Gramática histórica da Língua Portuguesa*. Melhoramentos.
- Said Ali, M. (1965). *Gramática secundária da Língua Portuguesa*. Melhoramentos.
- Saint-Dizier, P. (2006). *Syntax and semantic of prepositions*. Springer.
- Talmy, L. (2005). The fundamental system of spatial schemas in language. In B. Hampe, & J. E. Grady (Ed.), *From perception to meaning: image schemas in Cognitive Linguistics* (pp. 199-234). Mouton de Gruyter.
- Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: spatial scenes, embodied meaning and cognition*. Cambridge University Press.
- Vilela, M., & Koch, I. V. (2001). *Gramática da Língua Portuguesa*. Almedina.

Orações interrogativas com *qual*: análise gramatical e variação português europeu / português brasileiro¹

Interrogative clauses with qual 'which': grammatical analysis, and variation between European and Brazilian Portuguese

Recebido: 09/03/2024

Aceite: 17/07/2024

Publicado: 23/12/2024

Telmo Móia

Universidade de Lisboa (Portugal)

Centro de Linguística da Universidade de Lisboa (Portugal)

telmomoia@gmail.com | tmoia@letras.ulisboa.pt

ABSTRACT: This paper discusses the distribution of the interrogative morpheme *qual* 'which' in direct and indirect questions, considering four main types of syntactic contexts: equative predications (*qual [ser NP]* 'which [be NP]'), partitive constructions (*qual de NP* 'which of NP'), constructions with adnominal *qual* (*qual N'* 'which N'), and constructions with *qual* combined with null nominal forms (*qual []_{N'/deNP}* 'which []_{N'/ofNP}'). The prevalence of each construction and of additional grammatical phenomena, such as sluicing, insertion of *é que* and omission of the second NP in equative constructions, is evaluated using data from extensive corpora of newspaper texts. Aspects of the variation between European Portuguese (EP) and Brazilian Portuguese (BP) are also explored in some detail; these involve mainly the use of adnominal *qual*, a form that is extremely frequent in BP and almost residual in EP, except in direct questions superficially coinciding with the interrogative constituent. The role of three grammatical factors that seem to favour the use of adnominal *qual* in contemporary EP is also assessed: presence of prepositions in the interrogative constituent, spelling out of alternative answers, and establishment of anaphoric links with preceding nominal expressions. The analyses take into account data from Portuguese literary texts from the 16th to the 20th centuries. The competition between the equivalent forms *qual N'* and *que N'* in Brazilian Portuguese is also briefly examined.

KEYWORDS: *Qual* ('which'); Interrogative pronouns; Interrogative constituents; Interrogative clauses; Determiners; Language variation.

RESUMO: Este artigo discute a distribuição do morfema interrogativo *qual* em interrogativas diretas e indiretas, considerando quatro grandes tipos de contextos sintáticos: predicções equativas (*qual [ser SN]*), construções partitivas (*qual de SN*), construções com *qual* adnominal (*qual N'*) e construções com *qual* associado a formas nominais nulas (*qual []_{N'/deSN}*). Avalia-se, a partir de dados de extensos *corpora* de texto jornalístico, a prevalência de cada contexto e de fenómenos gramaticais suplementares, como o truncamento de toda a frase após o constituinte interrogativo (*sluicing*), a inserção

¹ Este trabalho foi financiado com verbas do Projeto Estratégico do Centro de Linguística da Universidade de Lisboa, UIDB/00214/2020.

de *é que* e a omissão do segundo SN em construções equativas. Exploram-se ainda, com algum pormenor, aspetos da variação português europeu (PE) / português brasileiro (PB), que envolvem essencialmente o uso de *qual* adnominal, forma muito frequente em PB, mas quase residual em PE, a não ser em interrogativas diretas superficialmente coincidentes com o constituinte interrogativo. É ainda avaliado o papel de três fatores que parecem favorecer o uso de *qual* adnominal em PE contemporâneo: a presença de preposições no constituinte interrogativo, a explicitação de alternativas de resposta e o estabelecimento de ligações anafóricas a expressões nominais precedentes. Nas análises, têm-se em conta, suplementarmente, dados de texto literário português dos séculos XVI a XX. É ainda brevemente avaliada a competição entre as formas equivalentes *qual* N' e *que* N' em português brasileiro.

PALAVRAS-CHAVE: *Qual*; Pronomes interrogativos; Constituintes interrogativos; Orações interrogativas; Determinantes; Variação linguística.

Introdução

Este trabalho discute a distribuição e aspetos de variação intralinguística do morfema interrogativo *qual*. Tradicionalmente, este morfema é classificado – a par do seu equivalente *que*, exemplificado em (2) – como um pronome interrogativo adjetivo ou adjunto (cf. e.g. Silva Dias, 1894 [1876]: 34; Figueiredo & Ferreira, 1973 [1965]: 219)², por poder ocorrer adnominalmente como parte de um SN, na sequência [_{SN} *qual* N'] – cf. (1).

- (1) De [_{SN} qual cidade italiana] gostas mais, Florença ou Veneza?
- (2) De [_{SN} que cidade italiana] gostas mais, Florença ou Veneza?

O estatuto categorial deste morfema *qual* (e bem assim de *que*, em *que* N') é objeto de discussão na literatura. Os autores oscilam geralmente entre uma classificação como determinante (especificador) ou como adjetivo (modificador), registando-se grande variação terminológica.³ Não me ocuparei aqui desta

² Silva Dias (1894 [1876]) divide os pronomes em “substantivos” e “adjetivos” (consoante sejam formas pró-SN ou parte integrante de um SN maior, respetivamente); a Nomenclatura Gramatical Portuguesa, de 1967, divide os pronomes de forma afim, com as designações de “absolutos” e “adjuntos”, respetivamente.

³ Brito (2003: 464, 474) classifica *qual* como “especificador” e como “quantificador”. Barbosa, Santos & Veloso (2013: 2540) classificam-no como “determinante” e como “especificador-determinante”. A classificação de formas afins de outras línguas também evidencia variação terminológica: Payne & Huddleston (2002: 356, 397) denominam o morfema inglês *which* (em *which* N') como “interrogative determinative”; Grevisse (1993 [1936]: 925) denomina o morfema francês *quel* (em *quel* N') como

questão.

O morfema *qual* tem uma distribuição muito variada, que será considerada em pormenor na secção 1, ocorrendo frequentemente como genuíno pró-SN em predicacões equativas (numa das análises possíveis destas construções, que aqui adoto, assumindo a estrutura argumental SER (x_{SN} , y_{SN}) para o verbo *ser* de identidade, de que *qual* é um dos argumentos).

- (3) [_{SN} Qual] é a cidade italiana mais bonita, na tua opinião: Florença ou Veneza?

Considerarei que em frases como (3) não existe elipse nominal, sendo o morfema *qual* classificável, segundo a terminologia tradicional, como pronome substantivo ou absoluto. *Qual* não é tradicionalmente classificado como tal (cf. e.g. Silva Dias, 1894 [1876]: 34), mas na análise aqui adotada, tem de facto uma variante substantiva (pró-SN), que é a ilustrada em (3).

O morfema interrogativo *qual* tem especial interesse gramatical pela diversidade de contextos em que pode surgir e porque, no seu uso adnominal ilustrado em (1), documenta uma interessante área de variação entre o português europeu (PE) e o português brasileiro (PB) contemporâneos, insuficientemente caracterizada na literatura, na minha opinião, e que será explorada em pormenor na secção 3. Sobre esta variação, Barbosa, Santos & Veloso (2013: 2542) comentam: “no português do Brasil, é mais comum do que em português europeu a ocorrência de *qual* em construções não partitivas”; “em português europeu, *qual* tende a ocorrer mais frequentemente em contextos particulares”. De entre os contextos mencionados por estes autores, interessam aqui apenas dois, ilustrados em (4) e (5):

- (4) A - Traz-me aí os livros! B - Quais livros? (Barbosa, Santos & Veloso, 2013: 2542)
- (5) O teu pai deu-te qual livro? (Barbosa, Santos & Veloso, 2013: 2542)

“déterminant interrogatif”; Fava (1995: 90) denomina o morfema italiano *quale* (em *quale* N') como “aggettivo interrogativo”.

Sobre as construções com *qual* adnominal, Brito (2003: 464, n. 34) observa que “certos gramáticos consideram que *qual* não deve ser usado seguido de nome [*quais livros compraste?*] e preferem o seu emprego em frases interrogativas predicativas como em *Qual é o hotel?* ou em construções partitivas como *Qual dos senhores...?*”. Esta autora coloca pois a questão em termos de aceitabilidade ou gramaticalidade, não apenas de menor frequência, um aspeto a que voltarei adiante.

O presente trabalho considera centralmente dados obtidos em *corpora* de texto jornalístico (ou predominantemente jornalístico) contemporâneo, português e brasileiro, nomeadamente: (i) o CETEMPúblico, com cerca de 195 milhões de palavras, para o português europeu, e (ii) o NILC/São Carlos, com cerca de 34 milhões de palavras, para o português brasileiro.

1. Contextos de ocorrência do morfema interrogativo *qual* - uma tipologia de construções

O morfema interrogativo *qual* ocorre em diferentes tipos de construções.⁴ A classificação destes usos de *qual*, ou das suas contrapartidas noutras línguas românicas, apresenta elevada variação terminológica na literatura, que aqui ignorarei, utilizando designações próprias (inspiradas em diversas propostas). Creio que são de distinguir quatro tipos de morfema interrogativo *qual*, o primeiro ilustrando um uso de *qual* como pronome substantivo ou absoluto (i.e. pró-SN) e os restantes três usos de *qual* como pronome adjetivo ou adjunto (isto é, elementos constituintes de um SN maior):

⁴ *Qual* também ocorre com frequência noutras construções que não serão aqui consideradas, nomeadamente em exclamativas (e.g. *Qual calma, qual carapuça!*) e em comparativas *lato sensu*, expressando um valor de semelhança, não de grau, parafraseável por *como se fosse* (e.g. *ela vaticina o futuro qual sibila*).

Tipo A. *Qual* como genuíno pró-SN em predicacões equativas - [_{SN} *qual*](*ser* SN)

(6) Qual é a capital de Itália?

Tipo B. *Qual* em estruturas partitivas, isto é, aplicado a um SN preposicionado (com *de*) - [_{SN} *qual de* SN]

(7) Qual das cidades que visitaste achaste mais interessante?

Tipo C. *Qual* adnominal interior ao SN, isto é, aplicado a um constituinte N' explícito - [_{SN} *qual* N']

(8) Qual cidade achaste mais interessante, Florença ou Veneza?

Tipo D. *Qual* isolado no constituinte interrogativo (CI), isto é, ocorrendo superficialmente como pró-SN, mas associado a uma estrutura nominal elidida (teoricamente analisável como *de* SN, como no tipo B, ou como N', como no tipo C, e representada aqui, por facilidade, como N*) - [_{SN} *qual* [N*]]⁵

(9) Visitaste três cidades italianas. Qual achaste mais interessante?

Concentrar-me-ei neste trabalho na ocorrência de constituintes interrogativos com *qual* em posição inicial de frase, como nos exemplos acima (que é, aliás, a posição mais comum). Porém, estes quatro tipos de *qual* também podem aparecer em posições não iniciais, como nos exemplos seguintes, para cada um dos subtipos, respetivamente: *O vinho que queres que leve é qual?*; *Gostas mais de qual destes livros?*; *Aceitaste qual proposta?*; *Visitaste três cidades. Gostaste mais de qual?*

Cada um dos quatro tipos A a D pode ser subdividido em função de certos

⁵ Sobre a possibilidade de elipse de ambos os tipos de estruturas nominais, veja-se: (i) *metade* [*dos eleitores*]; *foi votar e metade* []; *ficou em casa*; (ii) *nesta secção de voto, trezentos* [*eleitores*]; *votaram no partido A e duzentos* []; *no partido B*. Sobre a ocorrência de diferentes tipos de determinantes e quantificadores nestas posições dentro do SN, cf. Peres (2013: 754-760).

parâmetros gramaticais. Passo a descrever e exemplificar brevemente os que interessarão mais aqui.

Para o grupo A, das predicções equativas, destaco um parâmetro subclassificatório: a realização ou não do verbo *ser* de identidade, separando duas construções, ambas extremamente frequentes: predicções equativas com verbo *ser* realizado – cf. (10) – e predicções equativas com verbo *ser* omissa – cf. (11).⁶

(10) O diretor perguntou [_{SN} quais] são [_{SN} os projetos que vão ser financiados].

(11) O diretor perguntou [_{SN} quais] [_{SN} os projetos que vão ser financiados].

Note-se que o segundo SN da predicção equativa (*os projetos que vão ser financiados*) também pode estar elidido, se houver um antecedente adequado. Esta elisão ocorre em estruturas tanto com o verbo *ser* expresso como com o verbo *ser* omissa, como nos seguintes dois exemplos, respetivamente.

(12) Vários projetos vão ser financiados. O diretor perguntou [_{SN} quais] são.

(13) Vários projetos vão ser financiados. O diretor perguntou [_{SN} quais].

Para os grupos B a D, considerarei dois parâmetros subclassificatórios: (i) presença ou não de preposição no constituinte interrogativo; note-se que nas predicções equativas (grupo A) o constituinte interrogativo não pode ser preposicionado, mas nas restantes predicções pode; (ii) realização ou omissão da estrutura frásica após o constituinte interrogativo (e.g. *De certeza que gostaste mais de um dos dois vinhos que nos deram a provar. [Não sei] de qual gostaste mais. vs.*

⁶ Não considerarei neste trabalho uma construção, discutivelmente analisável como uma variante de (11), em que se realiza apenas o segundo SN da predicção equativa, sem o morfema interrogativo *qual*: *o diretor {perguntou / não sabe / não disse} os projetos que vão ser financiados*. As estruturas dependentes dos verbos *perguntar*, *saber* e *dizer* em frases como esta podem ser analisadas como estruturas proposicionais sem verbo e são referidas na literatura como *concealed interrogatives* (Baker, 1968), “interrogativas encobertas” (cf. e.g. Júnior, 2019) ou “interrogativas reduzidas” (Peres & Mória, 1995: 92).

[*Não sei*] de *qual* ~~*gostaste mais.*~~); note-se que o constituinte interrogativo pode surgir isolado, como a única parte visível da oração interrogativa, na chamada construção de *sluicing*, ou truncamento (cf. e.g. Ross, 1969, Merchant, 2001; Suñer, 1999).⁷ A combinação dos dois parâmetros em causa gera quatro subgrupos dentro de B e C e três grupos dentro de D. Por simplificação, no Quadro 1 adiante, não registarei os dados obtidos para o primeiro parâmetro para os grupos B e D, já que, ao contrário do que acontece no grupo C, ele não parece ser particularmente relevante. Vejamos cada caso separadamente.

Nas construções do grupo B, *qual* aplica-se a um SN preposicionado em construções partitivas, isto é, na combinação *qual de* SN.⁸ O constituinte interrogativo pode ser não preposicionado ou preposicionado, como nos dois exemplos a seguir, respetivamente:

(14) O diretor perguntou [_{SN} quais dos projetos] vão ser financiados.

(15) O diretor perguntou [_{SP} a quais dos projetos] iria ser atribuído financiamento.

Em qualquer dos dois casos, a oração interrogativa pode ser realizada apenas sob a forma de constituinte interrogativo, estando tudo o resto truncado:

(16) Vários projetos vão ser financiados. O diretor perguntou [_{SN} quais deles].

(17) Vai ser atribuído financiamento a vários projetos. O diretor perguntou [_{SP} a quais deles].

Nas construções do grupo C, *qual* aplica-se adnominalmente a um N' explícito, ocorrendo dentro do SN, na posição típica dos especificadores, em

⁷ O termo *sluicing* tem sido por vezes usado para referir qualquer forma de truncamento de toda a frase após o constituinte interrogativo, quer este constituinte seja simples (*quem, qual, onde,...*), como acontece porventura nas estruturas mais prototípicas, quer contenha elementos nominais (*que/qual pessoa, quantas pessoas,...*).

⁸ Ainda que não seja um quantificador, estando antes mais próximo de um determinante, *qual* tem, neste caso, uma distribuição sintática semelhante à de vários quantificadores - cf. e.g. *alguns [deles], cinco [deles], muitos [deles], metade [deles], parte [deles]*.

sequências com a forma *qual* N'.

(18) O diretor perguntou [_{SN} quais projetos] vão ser financiados. (#PE)

(19) O diretor perguntou [_{SP} a quais projetos] iria ser atribuído
financiamento.

Como já foi dito (cf. (1) vs. (2) acima, e ainda a secção 3.4 adiante), o morfema *qual* em *qual* N' é um equivalente sintático e semântico do morfema interrogativo *que* em *que* N'. Veremos melhor mais tarde que a distribuição deste morfema *qual* é fortemente condicionada em PE contemporâneo, muito mais do que em PB, um aspeto mencionado, mas pouco desenvolvido na literatura. O morfema *qual* adnominal sem truncamento da frase após o constituinte interrogativo - como em (18) e (19) - é muito comum em PB (podendo as estruturas preposicionadas ser sentidas como algo formais); aparentemente, foi mais comum em PE no passado (pelo menos, há bastantes abonações dele em autores clássicos portugueses - cf. secção 3.3), mas caiu praticamente em desuso em PE contemporâneo, principalmente quando o constituinte interrogativo não é preposicionado, como em (18), a ponto de essa frase poder soar pouco natural aos ouvidos dos falantes portugueses.

Nas construções do tipo C, a oração interrogativa também pode ser realizada apenas sob a forma de constituinte interrogativo, estando tudo o resto truncado. Vejam-se exemplos, em (20) e (21), com interrogativas indiretas (alíneas a) e diretas (alíneas b). Curiosamente, o uso de *qual* adnominal com truncamento da frase é bastante comum em PE contemporâneo, principalmente em interrogativas diretas com constituintes interrogativos não preposicionados (como em (20b)).

(20) a. A - Vários projetos vão ser financiados. O diretor perguntou [_{SN}
quais projetos].

b. A - Vários projetos vão ser financiados. B - Quais projetos?

- (21) a. Vai ser atribuído financiamento a vários projetos. O diretor perguntou [_{SP} a quais projetos].
b. A - Vai ser atribuído financiamento a vários projetos. B - A quais projetos?

Finalmente, nas construções do grupo D, *qual* aparece isolado no constituinte interrogativo em predicacões tipicamente não equativas, isto é, ocorre superficialmente como pró-SN. Porém, o morfema interrogativo aplica-se (plausivelmente) a uma expressão nominal elíptica, ligada anaforicamente a uma expressão nominal precedente, pelo que considerarei que é um pronome do tipo adjunto e não absoluto.⁹

- (22) Estava a falar-se dos projetos submetidos.
O diretor perguntou [_{SN} quais [_{N*}]] vão ser financiados.
(23) Estava a falar-se dos projetos submetidos.
O diretor perguntou [_{SP} a quais [_{N*}]] vai ser atribuído financiamento.

Como já foi referido brevemente, estas construções admitem teoricamente duas análises subjacentes distintas do constituinte interrogativo: com elipse de *de* SN, com antecedente *os projetos financiados* em (22)-(23), ou com elipse de N', com antecedente *projetos financiados* em (22)-(23). Consoante a análise, estas construções do grupo D seriam análogas às do grupo B ou do grupo C, respetivamente. Não tentarei aqui determinar se uma das análises é gramaticalmente mais bem motivada que a outra, representando a categoria elidida, por simplificação, como N*.¹⁰

Nestas construções, também pode haver truncamento de toda a sequência frásica pós-constituinte interrogativo. Ilustra-se a seguir o caso com constituinte interrogativo preposicionado, isto é, paralelo a (23):

⁹ A expressão nominal não realizada pode também ser interpretada deiticamente. Considere-se um contexto em que alguém mostra dois quadros a um amigo, perguntando: *De qual gostas mais?*, ou *Sabes qual é o meu preferido?*

¹⁰ Neves (1999: 539) ou Barbosa, Santos & Veloso (2013: 2542) defendem uma análise destas construções com elipse de N'.

(24) Vai ser atribuído financiamento a vários projetos. O diretor perguntou [SP a quais].

Teoricamente, *qual* isolado no constituinte interrogativo também pode ser associado a constituintes interrogativos não preposicionados, mas nesse caso a construção é indistinta de uma predicação equativa com omissão do verbo *ser* de identidade e do segundo SN, como a de (13) acima, repetida a seguir:¹¹

(25) Vários projetos vão ser financiados. O diretor perguntou [SN quais].

Nas contagens por grupos e subgrupos que realizei na secção 2, apresentadas no Quadro 1, contabilizei as construções do tipo de (25) – aliás, bastante comuns – como variantes da construção equativa, isto é, dentro do grupo A (com *ser* omissa) e não dentro do grupo D.

2. A distribuição contemporânea do morfema interrogativo *qual* – dados de *corpora* de texto jornalístico português e brasileiro

2.1. Aspectos gerais e metodologia

Para apurar dados sobre a frequência dos diferentes tipos de *qual* interrogativo, pesquisei sistematicamente *corpora* de texto jornalístico (ou predominantemente jornalístico) português (CETEMPúblico) e brasileiro (NILC/São Carlos). De forma a ter um volume de dados manejável, mas ainda assim suficientemente representativo, fiz duas pesquisas sistemáticas com limitações nas *queries*, que captam amostras representativas de interrogativas indiretas e de interrogativas diretas: Pesquisa 1 e Pesquisa 2, adiante. O número total de interrogativas obtidas (e analisadas) é 12.169 para o PE eu e 2.886 para o PB. Os resultados são apresentados no Quadro 1. Note-se que, nesse quadro, os valores apresentados entre parênteses curvos correspondem às estruturas com

¹¹ Note-se que havendo preposição, como em (24), não há coincidência possível com predicções equativas, visto que estas não ocorrem com constituintes interrogativos preposicionados.

truncamento da parte pós-constituente interrogativo, ou *sluicing*, sinalizadas como "(+ CI isolado)". As pesquisas realizadas foram:

Pesquisa 1 (conjunto de *queries* que captam uma amostra extensa de interrogativas diretas). Sequências começadas por *qual/quais* com maiúsculas ou com preposições escritas com maiúscula seguidas imediatamente de *qual/quais*: várias subpesquisas de "Qual|Quais"¹² e "A|Ante|Após|Até|Com|Contra|De|Desde|Em|Entre|Para|Perante|Por|Sem|Sob|Sobre" "qual|quais". Todos os resultados foram analisados e categorizados individualmente de modo a eliminar estruturas irrelevantes (e.g. exclamativas).

Pesquisa 2 (conjunto de *queries* que captam uma amostra extensa de interrogativas indiretas). Sequências com três dos verbos que mais comumente selecionam interrogativas indiretas (*perguntar*, *saber* e *dizer*) imediatamente adjacentes ao morfema interrogativo *qual/quais* ou separados dele por uma preposição:

[lema="perguntar|saber|dizer"] "qual|quais";

[lema="perguntar|saber|dizer"] [pos="PRP.*"] "qual|quais".

Obteve-se um total de 4605 registos relevantes: 3.874 de PE e 731 de PB. Todos os resultados obtidos foram analisados e categorizados individualmente, com uma exceção: no caso de *saber* imediatamente adjacente a *qual/quais*, no CETEMPúblico, o número de ocorrências é muito elevado (3.074); nesse caso, e apenas nele, foi considerada uma amostra (os primeiros 500 resultados), sendo os valores projetados proporcionalmente no total; os valores obtidos por projeção são sinalizados no Quadro 1 com o símbolo "≈".

2.2. Resultados das pesquisas em corpora

¹² A pesquisa simples "Qual|Quais" gera um volume de resultados difícil de manejar. Por isso, optei por subpesquisas direcionadas aos tipos relevantes de construções: para o tipo A com verbo *ser* realizado, "Qual|Quais" [lema="ser|estar|dever|poder|ir|vir|ter|haver"]; para o tipo A com verbo *ser* omissivo, "Qual|Quais" "o|a|os|as|sua|seu|suas|seus" (+ pesquisas excluindo elementos à direita do morfema interrogativo para captar as formas isoladas *Qual? Quais?*); para o tipo B, "Qual|Quais" "de.*|dos|das"; para o tipo C, "Qual|Quais" [pos="N.*" & pos!="NUM.*"] (+ pesquisa "Qual|Quais" "outr.*"); para o tipo D, "Qual|Quais" [pos="V.*" & lema!="ser"].

Os resultados das pesquisas realizadas estão sumariados no Quadro 1.

QUADRO 1 - Distribuição do morfema interrogativo *qual* por subtipos de construções em interrogativas diretas iniciadas por maiúscula e interrogativas indiretas imediatamente adjacentes a *perguntar*, *saber* e *dizer*, nos corpora CETEMPúblico (PE) e NILC/São Carlos (PB)

tipo de <i>qual</i>		pronome substantivo		pronome adjetivo				TOTAL
		A <i>qual</i> pró-SN (em construções equativas): [SN <i>qual</i>]		B <i>qual</i> em estruturas partitivas: [SN <i>qual de</i> SN] (+ CI isolado)	C <i>qual</i> adnominal interior ao SN: [SN <i>qual</i> N']		D <i>qual</i> isolado no constituinte interrogativo: [SN <i>qual</i> [N*]] (+ CI isolado)	
		<i>ser</i> realizado	<i>ser</i> omissivo (+ CI isolado)		CI _[-PREP] (+ CI isolado)	CI _[+PREP] (+ CI isolado)		
Interrogativas diretas <i>Qual</i> / <i>P qual... ?</i> (frases iniciadas por maiúsculas)	PE	5662	2077 (+214)	140 (+11)	4 (+144)	3 (+3)	31 (+6)	8295
		7953 (95,9%)		151 (1,8%)	154 (1,9%) 7 em CI não isolado (0,1% do total)		37 (0,4%)	
	PB	972	993 (+35)	48 (+6)	60 (+14)	17 (+5)	5 (+0)	2155
		2000 (92,8%)		54 (2,5%)	96 (4,5%) 77 em CI não isolado (3,6% do total)		5 (0,2%)	
Interrogativas indiretas <i>perguntar</i> / <i>saber</i> / <i>dizer</i> (P) <i>qual</i> (imediatamente adjacentes)	PE	≈1892	≈1658 (+≈137)	≈142 (+≈2)	5 (+3)	2 (+0)	≈34 (+≈1)	3874
		≈3685 (95,1%)		≈144 (3,7%)	10 (0,3%) 7 em CI não isolado (0,2% do total)		≈35 (0,9%)	
	PB	350	218 (+22)	18 (+0)	92 (+4)	16 (+4)	7 (+0)	731
		590 (80,7%)		18 (2,5%)	116 (15,9%) 108 em CI não isolado (14,8% do total)		7 (0,9%)	
TOTAL	PE	11631 (95,6%)		295 (2,4%)	171 (1,4%) 14 em CI não isolado (0,1% do total)		72 (0,6%)	12169
	PB	2590 (89,7%)		72 (2,5%)	212 (7,4%) 185 em CI não isolado (6,4% do total)		12 (0,4%)	2886

(Fonte: elaborado pelo autor)

Os dados sobre as construções dos grupos A, B e D serão analisados seguidamente. Discutirei primeiro os grupos B e D, que são mais simples, e depois A, que é ligeiramente mais complexo. Farei ainda algumas observações sobre factos que não estão registados no Quadro 1, nomeadamente a taxa de inserção de *é que* em interrogativas com *qual*. Os dados do grupo C - que são os mais interessantes e aqueles em que se observa maior variação PE/PB - serão analisados numa secção autónoma (secção 3).

Quanto às construções do grupo B (*qual* partitivo), apenas observarei que elas são usadas tanto em PE como em PB, com taxas de frequências relativamente modestas, globalmente quase idênticas nas duas variedades, em torno de 2,5% (do total de interrogativas com *qual*). Em PE, observa-se uma ligeira diferença de frequência nas interrogativas indiretas vs. interrogativas diretas (3,7% vs. 1,8%), que não se observa no PB (2,5% em ambos os tipos de interrogativas).

Quanto às construções do grupo D (*qual* isolado no constituinte interrogativo), apenas observarei que elas são usadas, tanto em PE como em PB, com taxas de frequências baixas, e também muito semelhantes nas duas variedades: em PE 0,6% e em PB 0,4% (do total de interrogativas com *qual*). Observa-se, em ambas as variedades uma pequena diferença de frequência nas interrogativas indiretas vs. interrogativas diretas (0,9% vs. 0,2-0,4%).¹³

Quanto às construções do grupo A (predicações equativas, ou identificacionais), são sem dúvida aquelas em que o morfema interrogativo *qual* ocorre com maior frequência no português. No PE, as predicações equativas representam mais de 95% das construções com *qual* interrogativo, tanto nas interrogativas indiretas como diretas; já no PB, representam apenas cerca de 81% das interrogativas indiretas e de 93% das interrogativas diretas. Os valores mais baixos em PB devem-se ao facto - que explorarei melhor na secção 3 - de que no Brasil a construção de tipo C é bastante frequente (quase 16% das interrogativas indiretas e quase 5% das interrogativas diretas), ao passo que esta construção tem

¹³ Em conformidade com o que foi mencionado no final da secção 1, a propósito da frase (25), as orações reduzidas a constituintes interrogativos não preposicionados (*qual* ou *quais*) foram integradas no grupo A (onde o seu número está registado entre parênteses curvos). Se considerássemos também essas estruturas no grupo D, a taxa de prevalência deste grupo seria de 3,8% para o PE (4,4% nas interrogativas indiretas, 3,0% nas diretas) e 2,4% para o PB (4,0% na interrogativas indiretas, 1,9% nas diretas).

em PE um uso quase residual (< 0,5% e < 2%, respetivamente). Vejam-se exemplos de construções equativas em interrogativas diretas - cf. (26) - e indiretas - cf. (27) - sem e com omissão do verbo *ser* de identidade (nas alíneas *a* e *b*, respetivamente), no CETEMPúblico:

(26) a. "Quais são as suas modalidades desportivas favoritas?"

(CETEMPúblico, ext1497936-des-91b-2)

b. "Quais as suas influências no campo da pintura?"

(CETEMPúblico, ext552-clt-94a-2)

(27) a. "Se alguém me perguntasse qual é a melhor banda do concelho

(...), não ia dizer que fosse a de Almoçageme.

(CETEMPúblico, ext1094392-nd-91b-1)

b. "Quando me perguntaram qual a melhor localização para a ponte

(...), defendi que seria aquela que o ordenamento do território

ordenasse (...)." (CETEMPúblico, ext767923-eco-95a-1)

Nas pesquisas realizadas para o Quadro 1, a taxa de omissão do verbo copulativo em predicções equativas é globalmente bastante elevada, tanto no CETEMPúblico como no NILC/São Carlos, embora seja menor no primeiro *corpus*: 35% [PE] vs. 49% [PB], respetivamente (ou 33% [PE] vs. 48% [PB], se considerarmos só as predicções com os dois SNs realizados). Comparando as interrogativas indiretas com as diretas, observam-se curiosas diferenças entre PE e PB: no CETEMPúblico, a taxa de omissão de *ser* em interrogativas indiretas equativas é maior do que em interrogativas diretas equativas (49% vs. 29%, ou 47% vs. 27% se considerarmos só as predicções com os dois SNs realizados); no NILC/São Carlos, é o inverso: a taxa de omissão de *ser* em interrogativas indiretas equativas é menor que em interrogativas diretas equativas (41% vs. 51%, ou 39% vs. 51% se considerarmos só as predicções com os dois SNs realizados). A maior diferença PE/PB observa-se nas interrogativas diretas: enquanto que no CETEMPúblico só se usam interrogativas elípticas do tipo de (26b) em 27% dos casos, no NILC/São Carlos usam-se em 51% dos casos.

A taxa de omissão do segundo SN em predicções equativas, como em (12)

e (13) acima ([*vários projetos vão ser financiados*] o *diretor perguntou* {*quais são / quais*}), não está representada no Quadro 1. Por curiosidade, apurei-a, mas só em interrogativas indiretas: no CETEMPúblico é de 6,4% (5,7% nas estruturas com *ser* e 7,6% nas estruturas sem *ser*); no NILC/São Carlos é de 8,3% (7,7% nas estruturas com *ser* e 9,2% nas estruturas sem *ser*).

Para concluir esta subsecção, referirei dois aspetos gerais, envolvendo os quatro grupos de construções: truncamento após o constituinte interrogativo (*sluicing*) e inserção de *é que* (este último não representado no Quadro 1).

Globalmente, a percentagem de interrogativas que estão reduzidas ao constituinte interrogativo, no volume total de interrogativas com *qual*, é ligeiramente maior no CETEMPúblico do que no NILC/São Carlos: 4,3% (521 registos) [PE] vs. 3,1% (89 registos) [PB], respetivamente. Em PE, é ligeiramente mais baixa nas interrogativas indiretas que nas diretas (3,7% vs. 4,6%, respetivamente). Em PB, é o inverso: é ligeiramente mais alta nas interrogativas indiretas que nas diretas (4,1% vs. 2,7%, respetivamente). A questão do truncamento após o constituinte interrogativo tem especial relevância para as construções do grupo C e a variação PE/PB em interrogativas diretas, e voltarei a ela adiante, na secção 3.

Uma curiosidade que os dados de *corpora* revelam é que, em orações com *qual* e no tipo de registo jornalístico em apreço, a inserção da sequência *é que* é extremamente infrequente. No NILC/São Carlos, a inserção de *é que* em interrogativas com *qual* nunca ocorre (mesmo em pesquisas mais alargadas, sem restrições).¹⁴ No CETEMPúblico, a combinação de *qual* e *é que* é residual: em interrogativas indiretas há apenas 19 registos (0,5% do total de interrogativas desta classe), distribuídos de forma muito distinta pelos grupos A, B e D: A (5 ocorrências, 0,1% dos registos deste grupo), B (7 ocorrências, 4,9% dos registos deste grupo), D (7 ocorrências, 20% dos registos deste grupo); em interrogativas diretas, o uso é ainda mais infrequente - há apenas 18 registos (0,2% do total de interrogativas desta classe), distribuídos da seguinte forma pelos grupos A, B e D: A (8 ocorrências, 0,1% dos registos do grupo), B (5 ocorrências, 3,3% dos registos do

¹⁴ Mesmo no muito mais extenso Corpus Brasileiro (da Linguatca), com 908 milhões de palavras, que inclui registos muito informais, há apenas 13 excertos com *qual* interrogativo e inserção de *é que* (7 no grupo A, 3 no grupo B e 3 no grupo C).

grupo), D (5 ocorrências, 13,5% dos registos do grupo). É certamente uma questão de registo, já que na oralidade a construção *qual... é que* não é incomum (em PE) – e.g. *Não sei qual é que é o teu problema! Para de me chatear!* Na sua secção sobre interrogativas com inserção de *é que*, Barbosa, Santos & Veloso (2013: 2545-2547) observam, de forma genérica (não especificamente para o morfema *qual*) e sem aduzir dados: “a inserção da sequência *é que* nas orações interrogativas parciais (...) é cada vez mais frequente em português, exceção feita a registos formais, altamente cuidados”.

3. Qual adnominal - um caso particularmente interessante de variação PE/PB

3.1. Primeira análise dos dados do Quadro 1 relativos ao grupo C

No que respeita ao uso de *qual* adnominal, há dois parâmetros gramaticais que têm uma importância fundamental: o primeiro, e mais importante, é o truncamento ou não da frase após o constituinte interrogativo; o segundo é o estatuto de interrogativa indireta ou direta da oração. Alguns outros parâmetros, que também parecem importantes, mas têm menor peso, serão discutidos na secção 3.2.

Começamos pelas estruturas sem truncamento após o constituinte interrogativo. Os dados do Quadro 1 confirmam o uso quase residual em PE de *qual* N' sem esse truncamento, como em (18) e (19) acima (*o diretor perguntou {quais projetos vão ser financiados / a quais projetos iria ser atribuído financiamento}*). Isto acontece tanto em interrogativas indiretas (7 registos, 0,2% do total de interrogativas dessa classe) como em interrogativas diretas (7 registos, 0,1% do total de interrogativas dessa classe). Vejam-se dois exemplos (com constituintes interrogativos não preposicionados e sem retoma anafórica, que são a forma mais rara, e potencialmente sentida como menos natural pelos falantes portugueses):

- (28) “É preciso saber quais técnicas a Renamo usou para manter a sua posição.” (CETEMPúblico, ext126814-pol-92a-1)

- (29) “Quais informações são acrescentadas na edição portuguesa de «Banqueiros de Rapina» em relação ao «caso Totta?»”
(CETEMPúblico, ext1048412-eco-94b-1)

Este uso infrequente contrasta com a elevadíssima frequência das estruturas em causa em PB: 108 interrogativas indiretas (14,8% do total de interrogativas dessa classe) e 77 interrogativas diretas (3,6% do total de interrogativas dessa classe), nas pesquisas efetuadas no NILC-São Carlos. Tendo em conta que o NILC/São Carlos é um *corpus* 5,74 vezes mais pequeno que o CETEMPúblico, isto representa uma frequência da construção no PB 89 e 63 vezes maior que no PE para as interrogativas indiretas e diretas, respetivamente. Com efeito, interrogativas genuinamente abertas do tipo das de (30) e (31), comuns em PB, não seriam naturalmente feitas por jornalistas portugueses, que usam prototipicamente *que N'* nestes casos. Interrogativas subordinadas do tipo de (32) também não são comumente escritas por jornalistas portugueses.

- (30) “Folha - Quais poetas influenciaram seu trabalho?”
(NILC/São Carlos, par=Ilustrada--94a-1)
- (31) “Os bairros (...) estão localizados próximos (...) ao Terminal Marítimo da Petrobrás. Quais medidas de segurança são tomadas para evitar problemas?” (NILC/São Carlos, par=Cotidiano--94b-2)
- (32) “Perguntados sobre quais instruções o técnico dava aos jogadores, nenhum quis responder.” (NILC/São Carlos, par=Esporte--94b-1)

A pesquisa no NILC/São Carlos revelou ainda 2 registos de construções com presença da forma *que* equivalente a *é que*, que não existe em PE, e creio ser mais característica do PB oral informal (havendo no exemplo (33a), adicionalmente, supressão da preposição argumental, *com*):

- (33) a. “Folha - Qual estilista brasileiro que você mais se identifica?”
(NILC/São Carlos, par=Ilustrada--94b-2)

b. “Qual máquina que está sendo usada?”

(NILC/São Carlos, par=Especial--94a-1)

Vejam agora as estruturas com truncamento da sequência após o constituinte interrogativo. Os dados do Quadro 1 mostram que o uso de *qual N'* com esse truncamento, como em (20a) e (21a) acima ([*vários projetos vão ser financiados*] o diretor perguntou *quais projetos* / [*vai ser atribuído financiamento a vários projetos*] o diretor perguntou a *quais projetos*), parece ser relativamente raro em interrogativas indiretas, tanto no PE (3 registos, 0,1% do total de interrogativas dessa classe) como no PB (8 registos, 1,1% do total de interrogativas dessa classe; ainda assim, 15 vezes mais frequente do que em PE). Veja-se um exemplo desta construção, que envolve sempre retoma anafórica, no CETEMPúblico:

(34) “Veio a minha casa e disse-me que tinha de mudar o sítio do barco para perto de umas laranjeiras, mas não me disse quais laranjeiras.”

(CETEMPúblico, ext157315-soc-94a-1)

Já em interrogativas diretas com constituinte não preposicionado e truncamento do resto da frase, como em (20b) e (21b) acima ([A - *Vários projetos vão ser financiados*] B - *Quais projetos?* / [A - *Vai ser atribuído financiamento a vários projetos*] B - *A quais projetos?*), a situação é muito distinta. Neste caso, há no CETEMPúblico 147 registos (1,8% do total de interrogativas diretas com *qual*). Trata-se de facto de uma construção de elevada frequência em PE contemporâneo.

(35) “(...) o entrevistador perguntará (...): «O senhor (...) ouviu rádio, ontem, entre as 7 e as 10 horas da manhã?» Se a resposta for sim, perguntar-se-á: «Qual rádio?»” (CETEMPúblico, ext1396150-nd-91a-2)

Interessantemente, muitos dos registos desta construção do CETEMPúblico (não contei quantos) correspondem não a interrogativas puras (isto é, interrogativas genuinamente abertas, em que se espera uma resposta de instanciação da variável

representada por *qual*), mas a interrogativas retóricas ou afins.¹⁵ Vejam-se dois exemplos:

(36) “Crise? Qual crise? Os primeiros três meses deste ano, na Europa, foram dos melhores de sempre para os fabricantes de microcomputadores (...).” (CETEMPúblico, ext418755-clt-soc-93a-1)

(37) “«Qual dívida?», pergunta-lhe o governador, «você não tem dívida nenhuma.»” (CETEMPúblico, ext1472082-soc-95a-1)

No NILC/São Carlos, identificaram-se apenas 19 interrogativas diretas com constituinte não preposicionado *qual N'* e truncamento do resto da frase, o que, tendo em conta o tamanho dos *corpora*, dá uma frequência menor no PB para esta construção (1,3 vezes menor). Há ainda a registar que neste *corpus* praticamente não aparece o tipo de interrogativas retóricas ilustradas em (36)-(37), que são muito comuns em PE. Porventura, este facto explica parte da diferença de frequência; deixo a apreciação desta conjectura para investigação futura.

3.2. Para uma análise mais fina dos dados do Quadro 1 relativos ao grupo C em português europeu

Globalmente, o uso de *qual N'* sem truncamento após o constituinte interrogativo em PE é relativamente raro e as construções afiguram-se por vezes como pouco naturais, face às suas alternativas com *que N'*, como acontece com (28) e (29) acima. Porém, observando mais atentamente os dados, verifica-se que a sensação de falta de naturalidade nem sempre surge (aliás, como veremos, ela raramente surge nos exemplos concretos encontrados no CETEMPúblico). Assim, é importante fazer uma análise mais fina dos resultados, de modo a verificar quais os fatores gramaticais favorecedores da construção em PE.

Para tal, considereei a pesquisa de interrogativas diretas realizada para o Quadro 1, que é muito abrangente, e alarguei a pesquisa das interrogativas

¹⁵ Cf. Barbosa, Santos & Veloso (2013: 2547-2551), para uma distinção entre “interrogativas puras” e “interrogativas não puras” (de quatro classes: “eco”, “sem resposta”, “retóricas” e “de surpresa”).

indiretas, do seguinte modo:¹⁶

Pesquisa 2' (conjunto de *queries* que captam uma amostra extensa de interrogativas indiretas; pesquisas com menos restrições que a Pesquisa 2, referida na secção 2.1). Sequências com quaisquer verbos, nomes ou adjetivos imediatamente seguidos do morfema interrogativo *qual* - ou separados dele por uma preposição - imediatamente seguido por um nome: [pos="V.*|ADJ.*|N.*"] "qual|quais" [pos="N.*" & pos!="NUM.*"]; [pos="V.*|ADJ.*|N.*"] [pos="PRP.*"] "qual|quais" [pos="N.*" & pos!="NUM.*"].

Esta pesquisa alargada gerou um total de 14 registos de *qual N'* sem truncamento após o constituinte interrogativo, isto é, mais 7 que os encontrados na Pesquisa 2, o que constitui um número extremamente baixo de ocorrências. Nesta secção, farei uma análise das 21 interrogativas com *qual N'* sem truncamento após o constituinte interrogativo (14 indiretas e 7 diretas) encontradas nas Pesquisas 1 e 2', no CETEMPúblico.

O que a observação dos exemplos mostra é que é necessário ter em conta certos parâmetros adicionais, já que parece haver fatores favorecedores do uso de *qual N'* na variedade europeia, que tornam as construções naturais mesmo em PE. Referirei aqui três fatores que se destacam. Só um deles está contemplado no Quadro 1 e começo por esse.

Fator 1. Presença de preposições no constituinte interrogativo.

A meu ver, a presença de uma preposição no constituinte interrogativo torna as frases com *qual N'* sem truncamento após esse constituinte ligeiramente mais naturais que as contrapartidas sem preposição. Sintomaticamente, dos 21 registos desta construção no CETEMPúblico, 9 (42%) têm constituintes interrogativos preposicionados.

¹⁶ Esta pesquisa capta certamente uma percentagem muito elevada do total das construções existentes no *corpus*, só não captando as que têm elementos intercalados, e.g. *perguntar [a alguém] qual N'*; *saber [talvez] qual N'*.

- (38) “Mas se me perguntar se já tenho assegurado um jogador de nível mundial para a posição do Ricardo, isso é verdade. Só não sei em qual clube vai jogar.” (CETEMPúblico, ext666772-des-91b-1)
- (39) “Por quais regiões da antiga Jugoslávia andou?” (CETEMPúblico, ext950404-nd-95b-1)

Fator 2. Explicitação de alternativas de resposta (“interrogativas alternativas”)

Huddleston (2002: 867ss.) propõe uma classificação tripartida das orações interrogativas, que integra, além das “interrogativas polares” - de resposta fechada: sim, não (*Is it breathing?*) - e das “interrogativas variáveis” - de resposta genuinamente aberta (*Why isn't it moving?*), uma terceira classe, de “interrogativas alternativas” - de resposta condicionada, com explicitação de um conjunto de alternativas na própria interrogativa: *Is it alive or dead?* (p. 867), *Would you like to meet in the morning, the afternoon, or the evening?* (p. 868). Os exemplos de Huddleston (2002) são todos sem elementos pronominais, mas as interrogativas pronominais podem igualmente explicitar alternativas. É o que acontece com *qual*: *qual das datas preferes, dia 21 ou dia 23?* (*qual* partitivo); *qual é a origem do problema, fungos ou bactérias?* (*qual* em predicções equativas). Ora, curiosamente, *qual* adnominal funciona com plena naturalidade, creio, em interrogativas alternativas em PE contemporâneo:

- (40) “Inquirido por qual banda optaria - Heróis ou Madredeus - se voltasse à estaca zero, responde: «Depende do ano (...).»” (CETEMPúblico, ext1553333-nd-91b-2)
- (41) “Qual adversário teme mais: Carlos Pimenta ou Ferreira do Amaral?” (CETEMPúblico, ext1328404-soc-92a-2)

Dos 21 registos de *qual* N' encontrados no CETEMPúblico, 2 são deste tipo, que não é muito frequente.

Fator 3. Presença de elementos nominais que estabelecem cadeias anafóricas com (*qual*) N'.

Nos exemplos portugueses (28)-(29) e brasileiros (30)-(32), há interrogativas genuinamente abertas, em que - adicionalmente - o constituinte interrogativo não remete para nenhum elemento nominal já previamente mencionado no discurso. Este é o tipo que creio causar maior estranheza em PE. Quando há uma expressão nominal na frase precedente que estabelece uma cadeia anafórica (mais direta - cf. (42)-(43) - ou mais indireta - cf. (44)) com a parte nominal expressa no constituinte interrogativo, a naturalidade aumenta.

(42) "As injúrias à autoridade não contam porque a autoridade não soube explicar qual autoridade foi injuriada."

(CETEMPúblico, ext710243-soc-91b-2)

(43) "A senhora abre a porta e atira o insulto: «Vão mas é para a vossa terra (...)!» Os jovens reagem aos gritos: Os jovens reagem aos gritos: «Qual terra é essa?»" (CETEMPúblico, ext749640-soc-93b-1)

(44) "Não havia dia que a senhora não visitasse o cemitério (...). Donalena (...) desomenageava a morte. Como? Ela não sabia qual campa devia honrar. Cada vez se joelhava numa diferente."

(CETEMPúblico, ext1518852-nd-94b-1)

Dos 21 registos de *qual* N' sem truncamento após o constituinte interrogativo encontrados no CETEMPúblico, pelo menos 12 integram este tipo de ligação anafórica. Em 3 destes registos - como que combinando os fatores 2 e 3 -, a parte nominal que cria a relação anafórica contém um conjunto de alternativas (podendo a estrutura ser, neste caso, facilmente parafraseável por uma construção partitiva, *qual de SN*):

(45) "A ideia é (...) formar vários grupos (...) e administrar metadona, heroína ou morfina de formas diferentes. Algumas vezes o paciente sabe qual produto recebe, outras não."

(CETEMPúblico, ext312439-soc-93b-1)

Sem nenhum dos fatores acima referidos, o número de ocorrências para as pesquisas muito abrangentes feitas no CETEMPúblico, um *corpus* de quase 200 milhões de palavras, é absolutamente residual, consistindo essencialmente nos exemplos (28) e (29).¹⁷

3.3. Dados do *corpus* Vercial

A compreensão plena dos aspetos gramaticais em discussão ganhará certamente com uma investigação aturada sobre mudança linguística, que não é possível realizar aqui. Porventura, algumas diferenças sinalizam evoluções divergentes das duas variedades da língua, a portuguesa e a brasileira. Não sendo possível realizar tal investigação, pareceu-me, no entanto, interessante comparar os dados apurados no CETEMPúblico com os usos documentados em texto literário português, mais antigo. Realizei no *corpus* Vercial (com cerca de 14 milhões de palavras de texto literário português, dos séculos XVI a XX) também a Pesquisa 1 e a Pesquisa 2'. O resultado foi 84 interrogativas com *qual* N' sem truncamento após o constituinte interrogativo: 48 indiretas e 36 diretas. Destas, 21 (25%) têm constituintes interrogativos preposicionados (10 indiretas e 11 diretas), 5 são interrogativas alternativas (1 indireta e 4 diretas), como as de (40)-(41) - cf. (46) -, e muitas têm antecedentes nominais e ligações anafóricas, como as de (42)-(43). Ou seja, os fatores 1 a 3 referidos na secção 3.2 parecem ser pertinentes.

¹⁷ Há um outro, em discurso direto, que parece não representar português europeu padrão, mas porventura português brasileiro: "E é isto, mano. Me explique, caso lhe chegue o entendimento. Eu não sei *qual pensamento hei-de escolher*." (CETEMPúblico, ext1312613-clt-94a-2).

- (46) "Qual refúgio lhe oferecerá a religião; refúgio imediato, sólido, esperançoso?" (Alexandre Herculano, *O Pároco de Aldeia*, 1851)

Porém, sobejam ainda pelo menos uma trintena de exemplos de interrogativas (indiretas - cf. (47) - e diretas - cf. (48)-(49)) que documentam a sintaxe que parece ter caído em desuso em Portugal e ter sido mantida no Brasil. Muitas das interrogativas são porém retóricas (cf. (49)), um fator cujo peso não avaliarei aqui, mas que interessará aquilatar em trabalhos futuros.

- (47) "Era necessário que as monjas soubessem qual futuro as aguardava."
(Alexandre Herculano, *Eurico, o Presbítero*, 1844)
- (48) "- Que novas nos trazes, Velido? Qual caminho seguem os árabes?"
(Alexandre Herculano, *Eurico o Presbítero*, 1844)
- (49) "(...) quem saberá contar as tuas horas excruciantes?
Quais almas descerão do teu Calvário com o segredo dos teus suplícios?!" (Camilo Castelo Branco, *Noites de Lamego*, 1872)

3.4. A competição entre *qual* adnominal e *que* adnominal

Na generalidade dos contextos, *que* N' pode ser utilizado como um equivalente de *qual* N',¹⁸ sendo a opção pela primeira forma praticamente a única utilizada em português europeu contemporâneo em contextos com a frase interrogativa não truncada.

Para avaliar a prevalência em PB de *que* N' vs. *qual* N', fiz no NILC-São Carlos duas subpesquisas das pesquisas realizadas para o Quadro 1.¹⁹ Os resultados

¹⁸ Já *qual* N' nem sempre pode ser utilizado como um equivalente de *que* N'. Isso acontece, por exemplo, quando *que* opera como um quantificador de medição - cf. Brito, 2003: 464, n. 33, sobre a frase *que vinho trago?*, que equivale, numa das suas leituras, não parafraseável com *qual*, a *que quantidade de vinho trago?* ou *quanto vinho trago?* Ignorarei aqui os valores claramente quantitativos (não identificativos) de *que* N'.

¹⁹ Subpesquisa 1: interrogativas diretas com morfema interrogativo escrito com maiúscula e com constituinte interrogativo não preposicionado - 74 interrogativas com *qual* N' vs. 703 interrogativas com *que* N' (90%). Subpesquisa 2: interrogativas indiretas imediatamente adjacentes a *perguntar*, *saber* ou *dizer* com constituinte interrogativo preposicionado (e ainda com constituinte interrogativo não preposicionado, no caso de *perguntar*) - 26 interrogativas com *qual* N' vs. 204 interrogativas com *que* N' (89%).

revelaram uma proporção de 90 % de *que* N' (907 registos) e 10% de *qual* N' (100 registos). Ou seja, apesar de a construção com *qual* N' se manter muito comum, a construção equivalente com *que* N' é claramente predominante em PB contemporâneo. Observe-se a competição de formas no NILC/São Carlos:

(50) “- Quais políticas o sr. adotará para a pequena empresa, caso seja eleito presidente?” (NILC/São Carlos, par=Brasil--94b-1)

(51) “- Que alianças o sr. fará para o segundo turno?” (NILC/São Carlos, par=Especial--94a-2)

A análise dos dados destas duas subpesquisas revela ainda uma clara diferença entre PE e PB quanto à inserção de *é que*. No NILC/São Carlos, observa-se uma quase ausência de inserção de *é que* também nas construções com *que* N': das 907 interrogativas com *que* N' obtidas nas duas subpesquisas referidas, apenas 1 usa *é que* (e 1 a sua forma truncada *que*, do tipo de (33) acima).²⁰ Ou seja, a situação não é muito diferente da observada com *qual* N', onde a inserção de *é que* nunca ocorre. Já no CETEMPúblico, considerando as mesmas subpesquisas, há uma taxa de inserção de *é que* em interrogativas com *que* N' de 7,1% (115 registos) para as interrogativas indiretas e 4,7% (≈171 registos) para as interrogativas diretas (7,9% e 5,8%, respetivamente, se considerarmos apenas as interrogativas sem truncamento após o constituinte interrogativo).²¹ Observa-se pois que, no CETEMPúblico, a taxa de inserção de *é que* em interrogativas com *que* N' é significativamente maior que a observada nas interrogativas com *qual* nas suas diversas ocorrências (*qual é que ser SN*, *qual de SN é que*, *qual [N*] é que*), que é próxima de 0,5% nesse mesmo corpus (e *é*, portanto, muito infrequente no registo jornalístico em apreço).

²⁰ Mesmo no muito mais extenso Corpus Brasileiro (da Linguateca), com 908 milhões de palavras e muitos registos informais, só há cerca de 32 registos de interrogativas com *que* N' e inserção de *é que*.

²¹ No caso das interrogativas diretas, valor projetado a partir da análise dos primeiros 500 resultados da pesquisa "Que" [pos="N.*" & pos!="NUM.*"] (que gera 6108 resultados, incluindo relevantes e não relevantes).

Conclusões

Neste trabalho, foi observada a distribuição de *qual* interrogativo em quatro grandes tipos de contextos. Verificou-se uma forte predominância do uso deste morfema em predicções equativas, em ambas as variedades da língua: 96% e 90% dos registos de *qual* interrogativo no CETEMPúblico e no NILC/São Carlos, respetivamente. Das predicções equativas, cerca de um terço no CETEMPúblico e de metade no NILC/São Carlos apresentam omissão do verbo *ser*.

Observou-se ainda uma diferença significativa entre o PE e o PB no uso de *qual* adnominal, forma equivalente a - e em competição com - *que* adnominal. No PB documentado no NILC/São Carlos, a forma *qual* N' é escolhida em 10% dos casos e *que* N' nos restantes 90% dos casos; *qual* N' representa 16% dos usos de *qual* em interrogativas indiretas, a classe em que o seu uso é mais comum, e 5% dos usos de *qual* em interrogativas diretas. Nas pesquisas efetuadas, encontrou-se um total de 212 registos, 185 dos quais (87%) não envolvem truncamento da frase após o constituinte interrogativo.

No PE documentado no CETEMPúblico, a forma *qual* N' em estruturas sem truncamento é extremamente rara (0,1% do total de interrogativas com *qual*) e está por vezes associada a uma sensação de falta de naturalidade; há, porém, vários fatores que favorecem o uso de *qual* N' sem truncamento em PE: a presença de preposições no constituinte interrogativo, a explicitação de alternativas de resposta e o estabelecimento de ligações anafóricas a expressões nominais precedentes. As formas sem truncamento e sem nenhum dos três fatores favorecedores praticamente não ocorrem em PE (mas são bastante comuns em PB - cf. (30)-(32)). A forma *qual* N' só tem expressão significativa em PE contemporâneo em interrogativas diretas com truncamento após o constituinte interrogativo (principalmente com constituintes interrogativos não preposicionados), representando esta forma 1,3% do total de interrogativas diretas com *qual* no corpus.

Referências

- Baker, C. L. (1968). *Indirect questions in English*. [Unpublished PhD Dissertation]. University of Illinois.
- Barbosa, P., Santos, P., & Veloso R. (2013). Tipo de frase e força ilocutória. In E. B. P. Raposo, M. F. B. do Nascimento, M. A. C. da Mota, L. Segura, A. Mendes (Orgs.). *Gramática do Português* (pp. 2515-2586). Fundação Calouste Gulbenkian.
- Brito, A. M. (2003). Frases interrogativas. In M. H. M. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, F. Oliveira, M. Vigário, & A. Villalva, *Gramática da Língua Portuguesa* (5ª ed., pp. 460-479). Editorial Caminho.
- Fava, E. (1995). Il tipo interrogativo. In L. Renzi, G. Salvi & A. Cardinaletti (Orgs.), *Grande Grammatica Italiana di Consultazione III* (pp. 70-127). Il Mulino.
- Figueiredo, J. M. N. & Ferreira, A. G. (1973). *Compêndio de Gramática Portuguesa* (3ª ed). Porto Editora. (1.ª ed., Lisboa: Livraria Sá da Costa, 1965)
- Grevisse, M. (1993). *Le Bon Usage. Grammaire Française* (13ª ed.). Duculot. (1.ª ed., 1936).
- Huddleston, R. (2002). Clause type and illocutionary force. In R. Huddleston, & G. K. Pullum, *The Cambridge Grammar of the English Language* (pp. 851-945). Cambridge University Press.
- Júnior, P. M. (2019). Algumas considerações sobre a sintaxe das interrogativas indiretas encobertas do português do Brasil. *Caderno de Squibs: Temas em estudos formais da linguagem*, 5(2), 73-91.
- Merchant, J. (2001). *The syntax of silence: Sluicing, identity, and the theory of ellipsis*. Oxford University Press.
- Neves, M. H. M. (1999). *Gramática de usos do português*. Editora Unesp.
- Payne, J., & Huddleston, R. (2002). Nouns and Noun Phrases. In R. Huddleston & G. K. Pullum, *The Cambridge Grammar of the English Language* (pp. 323-523). Cambridge University Press.
- Peres, J. A., & Mória, T. (1995). *Áreas Críticas da Língua Portuguesa*. Editorial Caminho (2.ª ed., 2003).
- Peres, J. A. (2013). Semântica do sintagma nominal. In E. B. P. Raposo, M. F. B. do Nascimento, M. A. C. da Mota, L. Segura, A. Mendes (Orgs.). *Gramática do Português* (pp. 735-815). Fundação Calouste Gulbenkian.
- Ross, J. R. (1969). Guess who? In R. Binnick, A. Davison, G. Green & J. Morgan (Eds.), *CLS 5: Papers from the fifth regional meeting of the Chicago Linguistic Society* (pp. 252-286). Chicago Linguistic Society.
- Silva Dias, A. E. (1894). *Grammatica Portugueza Elementar* (9ª ed.). A Ferreira Machado & C.ª Editores. (1.ª edição, 1876)
- Suñer, M. (1999). La subordinación sustantiva: la interrogación indirecta. In I. Bosque, & V. Demonte (Dir.), *Gramática Descriptiva de la Lengua Española. Real Academia Española* (pp. 2149-2195). Espasa.

Corpora online consultados

CETEMPúblico 2.0 v. 12.1, <http://www.linguateca.pt/ACDC/> [último acesso: 01.03.2024]

Corpus Brasileiro v. 7.0, <http://www.linguateca.pt/ACDC/> [último acesso: 01.03.2024]

NILC-São Carlos v. 14.2, <http://www.linguateca.pt/ACDC/> [último acesso: 01.03.2024]

Vercial v. 16.7, <http://www.linguateca.pt/ACDC/> [último acesso: 01.03.2024]

Analyzing GPT-4 Misinterpretations of Russian Grammatical Constructions¹

Timofei Plotnikov

UiT The Arctic University of Norway (Norway)

timofei.plotnikov@uit.no

Recebido: 18/07/2024

Aceite: 02/11/2024

Publicado: 23/12/2024

ABSTRACT: Generative Pre-trained Transformers (GPTs), which are Large Language Models (LLMs) primarily trained on datasets dominated by English texts, can incorporate inaccuracies into their final outputs. The training data bias presents challenges for accurate performance in non-Roman script languages, such as Russian. Challenges are specifically expected with grammatical constructions in Russian which are colloquial, idiomatic, and non-compositional. Although the empirical investigation reveals that, overall, GPT-4 performs well with the majority of Russian grammatical constructions, it still encounters limitations with low-frequency constructions due to insufficient training data, a lack of context, and the influence of the English language.

KEYWORDS: Large Language Models; GPTs; Grammatical Constructions; Russian.

Introduction

Artificial Intelligence (AI) chatbots, which represent practical applications of LLMs, are seeing widespread use due to their advanced capabilities. These models can perform a variety of language-specific tasks such as improving, translating, interpreting text, responding to context, and engaging in meaningful conversations. One of the types of LLMs is GPT, predominantly trained on a vast amount of English data from the Internet (Brown *et al.*, 2020; Lai *et al.*, 2023; Ray, 2023; Wendler *et al.*, 2024). The most renowned AI chatbot, ChatGPT, which implements GPT-4 at the moment, was launched in November 2022 and played a crucial role in advancing research into human-LLM interaction.

While LLMs exhibit human-like capabilities, they are still prone to errors. Assessing the accuracy of the responses from LLMs remains a complex and

¹ Special thanks for assistance with the Python coding and processing are dedicated to the Department of Computer Science, UiT The Arctic University of Norway.

insufficiently explored issue, especially in languages other than English (Papadimitriou *et al.*, 2022; Lai *et al.*, 2023; Ray, 2023; Hendy *et al.*, 2023; Wendler *et al.*, 2024; Liu, 2024). OpenAI (the company behind GPTs and ChatGPT) states² that GPTs are proficient in working with the English language but can struggle with other languages, especially those that use a non-Roman script. This is why investigating Russian, which is not only one of the most spoken languages in the world but also uses a Cyrillic script, is important for the human-LLM interaction research.

The present study specifically focuses on analyzing the performance of GPT-4, which is implemented in the AI chatbot ChatUiT (<https://chat.uit.no/>), on a dataset that includes 2,277 Russian grammatical constructions collected from the Russian Constructicon (<https://constructicon.github.io/russian/>). The Russian Constructicon is an open-access platform featuring a searchable database of Russian multiword grammatical constructions designed specifically for researchers and second language (L2) learners (Bast *et al.*, 2021). Considering the potential of AI chatbots that implement GPTs to assist researchers and L2 learners of Russian in language-specific tasks, the following research questions arise: *how well does GPT-4 interpret Russian grammatical constructions? What are the factors contributing to GPT-4's misinterpretations with certain Russian grammatical constructions?*

In order to address the research questions, the primary objective of this study is to conduct an experiment exploring language-specific issues that users can face while using AI chatbots for tasks involving Russian. The final goal is to demonstrate AI performance in languages other than English while acknowledging limitations, thereby contributing to the future development of LLMs.

This article consists of 6 sections. Section 1 introduces the topic. Section 2 provides the reader with the theoretical foundation relevant to the research: an overview of the performance of LLMs applicable to the current study, and an introduction to grammatical constructions and Russian Constructicon. Section 3 describes the experimental setting. In Section 4, the obtained results are interpreted and analyzed. Section 5 summarizes the entire investigation and suggests ideas

² <https://openai.com/index/chatgpt-can-now-see-hear-and-speak/> (21.03.2024).

based on the findings. Section 6 highlights the limitations of the study and proposes future areas of research.

1. Background

1.1. Previous research on GPTs and their performance

Generative pre-trained transformer is a type of large language models developed by OpenAI, based on a deep learning architecture for handling natural language processing (NLP) tasks, primarily for text generation (Zhu & Luo, 2022; Yenduri *et al.*, 2024). GPTs are initially pre-trained on a large dataset of diverse texts and can perform a broad range of language-related tasks, enabling GPTs to generate human-like texts based on a given input. The dataset includes a wide variety of content sourced from a broad spectrum of internet platforms, notably including materials from Wikipedia and Google Books (Brown *et al.*, 2020). GPTs are designed to discover statistical patterns between words and their contexts in the training dataset using a technique called autoregressive language modeling, which is a building block of GPTs (Yenduri *et al.*, 2024). Autoregressive language modeling attempts to predict the next word based exclusively on the previous words sequentially processing the context from left to right (Brown *et al.*, 2020; Zhu & Luo, 2022; Pannatier *et al.*, 2024). This technique allows GPTs to generate accurate responses based on statistical predictions. Statistical predictions rely on statistical patterns learned from the training dataset and the frequent occurrence of words, aiming to capture a broader context (Gong *et al.*, 2019; Yenduri *et al.*, 2024). GPTs heavily depend on contextual information provided in the input, specifically on semantically similar examples, where phrases and sentences, despite their differences, share similar meanings within a single text (Liu *et al.*, 2021). By providing more context for GPT models, their performance will be more accurate, coherent, and less biased (Yenduri *et al.*, 2024). Without sufficient context, GPTs tend to misinterpret users' requests. This can lead to weak and untruthful performance and predisposes GPTs to frequency/statistical bias, especially with non-frequent and rare words (Brown *et al.*, 2020; Dudy & Bedrick, 2020: 13).

LLMs, including GPTs, are predominantly trained on massive corpora of English text (Lai *et al.*, 2023; Ray, 2023; Wendler *et al.*, 2024). According to Brown *et al.* (2020: 6), GPT-3 predominantly implements the Common Crawl text corpus, which is 93% English by word count and 7% non-English; Russian, one of the most spoken languages in the world, comprises just 0.2% of this corpus. Beginning with GPT-3.5, OpenAI has not disclosed the details of the datasets used to train the GPT models. However, more advanced GPTs still retain a practical bias towards the English language in their outputs (Lai *et al.*, 2023). Since the training data in the text corpus is heavily disproportionate, LLMs exhibit decreased performance in languages other than English. This also results in English being employed as an internal pivot language in LLMs; inputs in languages other than English are initially translated into English, processed, and then translated back into the original language, which leads to misinterpretations and failure to convey the intended meaning accurately (Wendler *et al.*, 2024). The English-language-biased data in LLMs not only leads to grammatical inaccuracies by imposing English syntactic patterns but also results in the inappropriate application of lexicon to other languages, where it may be irrelevant or inapplicable (Papadimitriou *et al.*, 2022). Besides, the biased data may introduce unexpected elements in the output, such as punctuation insertions, which result in an incorrect linguistic representation of other languages (Hendy *et al.*, 2023).

GPTs can also struggle with the interpretation of abstract concepts and non-literal language, which includes idioms and sarcasm (Ray, 2023; Yenduri *et al.*, 2024). Responses from GPTs tend to be overly literal, leading to misunderstandings of the original meaning and the desired tone. This limitation can result in incorrect translations and interpretations of idiomatic expressions, which are deeply embedded with cultural values. Johnson *et al.* (2022) report that translations from Russian through GPTs often fail to convey cultural values due to the underrepresentation of Russian in the training dataset. In contrast, GPTs tend to reflect American values more effectively due to the dominance of data available mostly in American and, to a lesser extent, British English (Dodge *et al.*, 2021; Johnson *et al.*, 2022; Jackson, 2023: 21; Liu, 2024: 114); this reveals that GPTs are primarily pre-trained to represent cultural-linguistic concepts within the

Anglosphere. As a result, the prevalence of American English texts can sometimes be visible in outputs generated in another language with a significant change in embedded cultural values (Johnson *et al.*, 2022). Therefore, it is important to note that English serves as a pivot language not only lexically but also semantically (Wendler *et al.*, 2024), since the role of LLMs is not limited to straightforward translation only. LLMs may fail to accurately convey cultural values in another language, but the biased output can still appear somewhat grammatically correct and acceptable to non-experts. Furthermore, the official blog³ of OpenAI (the company behind GPTs) cautions users that the model's outputs may not always be entirely accurate, since it is not an expert and does not take full responsibility for the provided content - only human experts can determine the relevance and accuracy of the output.

According to Tan *et al.* (2023), the GPT LLM family can struggle to generate relevant and accurate outputs for more complex tasks. However, not only GPTs but also many other LLMs exhibit significant challenges when compared on tasks of varying complexity levels that require reasoning abilities such as causal understanding, logical deduction, and counterfactual reasoning; these cognitive processes are complex and often associated with human-like consciousness, which LLMs do not possess (Gemini Team, 2023: 23; Lai *et al.*, 2023). Even GPT-4, a more advanced version of the GPT LLM family, does not consistently show improved performance in tasks requiring high-level reasoning abilities. Consequently, the lack of the reasoning abilities is critical for the precise interpretation of user inputs. This limitation is evident when dealing with grammatical constructions in Russian, which often include idiomatic expressions that are not amenable to a logical, compositional analysis due to their complex pragmatic and cultural nuances (Janda *et al.*, 2018). This issue is compounded by the nature of Russian grammatical constructions which often exhibit non-compositionality, i.e., the meaning of a construction cannot be interpreted solely from its individual components (Janda *et al.*, 2020: 163; Rakhilina *et al.*, 2022). Besides, some of these constructions are primarily idiomatic and colloquial and cannot be successfully interpreted based on

³ <https://help.openai.com/en/articles/6783457-what-is-chatgpt> (21.03.2024).

statistical predictions alone. Therefore, the concept of grammatical constructions and their components is being elaborated further in more detail.

1.2. The concept of grammatical constructions and a constructicon

Grammatical constructions are defined as “learned pairings of form and function, including words and idioms as well as phrasal linguistic patterns” (Goldberg & Suttle, 2010: 469). Learned pairings of form and function include common and minor patterns. Examples of common patterns encompass passive, topicalization, and relative clauses; minor patterns cover words and idioms. While varying in their generality, size, and complexity, both patterns are still considered grammatical constructions. The main aspects of constructions primarily include their semantic attributes, information and discourse characteristics, as well as usage conditions (e.g., register and genre) (Fillmore *et al.*, 1988; Goldberg, 2006; Goldberg & Suttle, 2010).

The linguistic theory that focuses on grammatical constructions is Construction Grammar (CxG). One of the main products that CxG provides is a *constructicon* as it supports the main idea of this theory that language is constructed according to a set of grammatical rules to convey meaning (Fillmore *et al.*, 1988; Goldberg, 2006; Goldberg & Suttle, 2010; Hilpert, 2014; Lyngfelt *et al.*, 2018; Janda *et al.*, 2024). A constructicon is a large network of various grammatical constructions of a specific language, where the constructions are grouped into families, clusters, and networks according to their semantic and syntactic characteristics; grammatical constructions are typically demonstrated with examples from corpora and labeled according to the CEFR levels of language proficiency (Hilpert, 2014; Lyngfelt *et al.*, 2018; Janda *et al.*, 2020). In practice, a constructicon represents a distinctive dictionary of grammatical constructions containing thorough annotations. The annotations of constructions usually include definitions, valid lexical references from phraseological/idiomatic dictionaries, and grammar references explaining the syntactic/semantic types of the construction, as well as usage labels (Janda *et al.*, 2020: 163).

The constructicon approach to constructions can be conceptualized as a complex grammar network model in which, according to Diessel (2019), different linguistic elements are linked and related. Within this theoretical model, constructions exhibit taxonomic and horizontal relations. Taxonomic relations are schemas in which components of constructions are generalized and categorized in order to be reused for the production of novel instances, while horizontal relations refer to how constructions are interconnected and used (Diessel, 2023). Thus, one of the types of grammatical constructions that can be represented in a constructicon contains a variable slot and a fixed part (anchor). The variable (open, non-fixed) slot can be filled with a set of lexemes (fillers) which represent words and phrases that are suitable for the grammatical construction (Janda *et al.*, 2020). Meanwhile, the fixed part, which is partially schematic, has a strict structure (e.g., word order and/or specific words) that is duplicated precisely and remains unchanged due to lexical and semantic constraints (Janda *et al.*, 2018; Bast *et al.*, 2021). Although the variable slot is “open”, it can still be subjected to certain lexical and semantic constraints. This idea is specifically implemented in *Russian Constructicon (RusCon)*, a one-of-a-kind searchable database of grammatical constructions with detailed annotations (Bast *et al.*, 2021):

ID1712 NP/VP za komp'juterom
Rabotat' *za* *komp'juterom*
 [work.INF behind computer.INS]
 'To work at the computer' (Bast *et al.*, 2021)

This example demonstrates that noun phrase/verb phrase (NP/VP) can vary and instead of *rabotat'* 'to work', there are also various lexemes such as *sidet'* 'to sit', *provodit' vremena* 'to spend time', *zasypat'* 'to fall asleep', *žizn' prohodit* 'life passes', *prosiživat'* 'to sit out', etc. that can fill the slot of NP/VP, while *za komp'juterom* 'at the computer' is the fixed part of the construction and cannot be changed.

The RusCon database, a collection of 2,227 entries, was created through a collaboration between UiT The Arctic University of Norway, the University of Gothenburg, and the National Research University Higher School of Economics by professional linguists, researchers, BA, MA, and PhD students (Janda *et al.*, 2018;

Bast *et al.*, 2021). Each entry includes one grammatical construction, featuring a fixed part and a variable slot. Entries also include a unique ID number (Janda *et al.*, 2024) that ranges from 1 to 2,227, a definition, examples (usually five additional examples featuring the same grammatical construction), CEFR levels (A1, A2, B1, B2, C1, C2), equivalents in other languages (e.g., translations of the grammatical construction into English), syntactic aspects (type, structure, function), dependency structure (the illustration of grammatical relationships between different parts of a sentence), communicative types, usage labels, extra comments, and scholarly references.

The 2,227 grammatical constructions of the RusCon are predominantly colloquial and underrepresented in written discourse, primarily manifesting idiomatic rather than schematic, transparent meanings (Janda *et al.*, 2020; Rakhilina *et al.*, 2022). Many of these constructions are not included in dictionaries or other reference works and have never been documented in scholarly literature before. The RusCon project attempted to primarily focus on high-frequency constructions. Therefore, phrases that potentially represent constructions but had fewer than 27 attestations in *the Russian National Corpus* were considered infrequent and usually excluded (Janda *et al.*, 2018). However, since many of the constructions are non-compositional, they typically had to be collected, translated, and described manually, due to their opaque forms and meanings not yet being comprehensible to computational systems. Additionally, there is an issue with the poor comparability of phraseological units across languages (Rakhilina *et al.*, 2022). Despite the existence of many bilingual dictionaries for major languages, bilingual phraseological dictionaries remain rare. Therefore, many items for RusCon were collected from idiomatic and phraseological dictionaries intended for native speakers; these dictionaries often lack detailed descriptions of the semantic and syntactic components of non-transparent expressions (Janda *et al.*, 2018). Such descriptions are crucial for non-native speakers and L2 learners to fully grasp the rich pragmatic and cultural implications of idiomatic constructions. Thereby, according to Janda *et al.* (2018: 167), RusCon is an important resource in addressing this need from the complementary perspectives of non-native language users.

In light of the extensive scope, non-compositionality, and detailed annotations of the RusCon database, the current investigation aims to utilize its 2,227 entries to provide insights into GPT-4's performance, considering that AI chatbots are used for translation and interpretation in various languages. Thus, the findings may be applicable to other languages, particularly those that use the Cyrillic script, and can potentially interest linguists, L2 learners, and NLP developers.

2. Data and methodology

In order to test how well GPT-4 interprets Russian grammatical constructions, a dataset of 2,227 examples was gathered from Russian Constructicon. Each of the 2,227 entries in the resource includes a grammatical construction with its own ID number followed by an illustrative phrase, e.g., ID3 (a) kak že NP-Nom? - *A kak že mama?* 'And what about mom?'; (a) kak že is the fixed part, while NP-Nom? is the variable slot. The fixed part of grammatical constructions is more critical for the focus of this study, as it potentially indicates that similar issues could occur in other examples featuring the same grammatical construction. Therefore, the results where the issue was located elsewhere rather than in the fixed part are less relevant to this investigation and are not examined in more detail.

All 2,227 illustrative phrases were subjected to testing through the AI chatbot ChatUiT, which implements the most advanced model of the GPT LLM family at the time the experiment was conducted: GPT-4. Prior to each example the question in English "What do you think of the following sentence?" was asked, followed by automated processing of the dataset using the Python programming language. Although the phrasing of the question might seem informal and clumsy, it was deliberately chosen to simulate a conversational manner similar to everyday human interactions, representing how non-scientists potentially interact with AI chatbots. Additionally, the task was assigned in English because RusCon is designed primarily for non-native speakers with annotations for those who are proficient in English but might not be fluent enough to understand linguistic descriptions in Russian. Thereby, conducting the experiment in English helps to reveal the accuracy and depth of GPT-4's translations and interpretations of Russian grammatical constructions.

RusCon also provides CEFR levels of language proficiency (A1, A2, B1, B2, C1, C2) for each entry, allowing the data to be divided across these six levels.

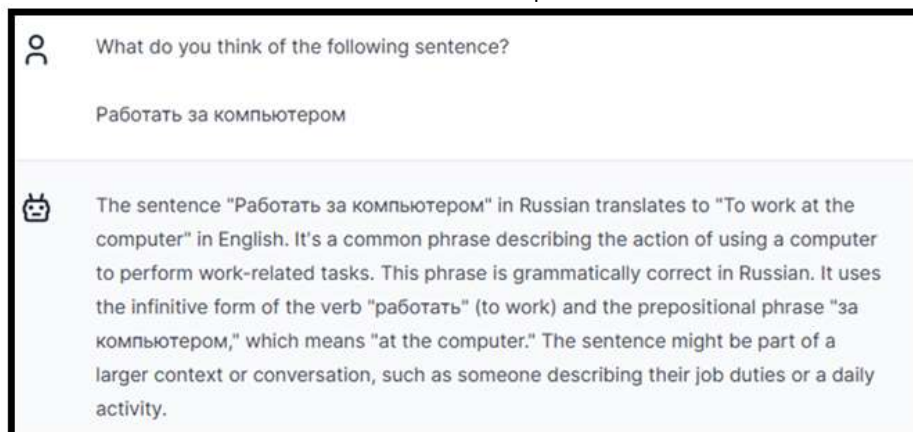
Most of the 2,227 illustrative phrases featuring a grammatical construction are accompanied by five additional examples in RusCon. The five additional examples of constructions that presented challenges for GPT-4 were also subjected to testing (more details in *Section 4*). The identical task was assigned to GPT-4, where the gathered data was automatically processed through Python, and the same question “What do you think of the following sentence?” was asked before each example.

Following the testing phase, I performed a manual evaluation of GPT-4 misinterpretations involving detailed comparative analyses with various elements of the annotations from RusCon. Further, data from the largest web corpora of Russian: *the Russian Web 2011 corpus (ruTenTen 2011)*, *the Russian National Corpus*, and *Google Books*, which are incorporated into the training dataset of GPTs, are being employed in this investigation to ascertain whether low frequency contributes to the misinterpretation of certain examples. Additional resources, such as dictionaries and relevant scholarly articles, are also being employed, given that annotations from RusCon sometimes do not provide sufficient information that is relevant to the analysis of misinterpretations.

3. Results

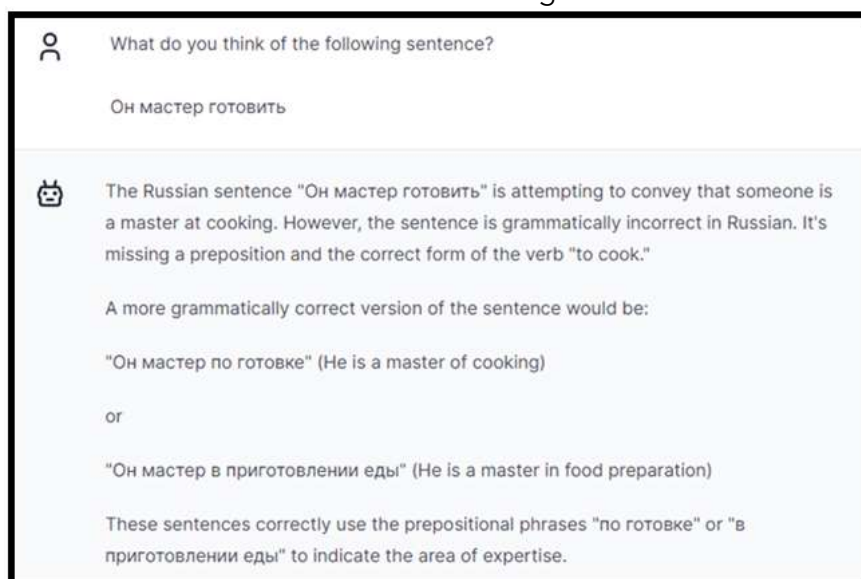
GPT-4 responses are predominantly heterogeneous across all 2,277 outputs, describing the meanings of illustrative phrases along with comments on their grammatical structures and contextual usages. Below, Figure 1 provides a GPT-4 response with a correct interpretation, while Figure 2 provides one with an incorrect interpretation.

FIGURE 1 - The response with a correct interpretation by GPT-4 in ChatUiT of the grammatical construction: ID1712 *NP/VP za komp'juterom - Rabotat' za komp'juterom* 'To work at the computer'



(Source: ChatUiT (2024), <https://chat.uit.no/>)

FIGURE 2 - The response with an incorrect interpretation by GPT-4 in ChatUiT of the grammatical construction: ID577 *NP-Nom Cop NP-Nom VP-Inf - Он мастер готовить* 'He is a master at cooking'



(Source: ChatUiT (2024), <https://chat.uit.no/>)

Out of 2,277 tested examples, GPT-4 generally performs well but struggles with the correct interpretation of 18 examples (GPT-4 outputs in text file format for all 2,277 inputs are available online⁴). All the 2,277 examples from RusCon adhere to standard Russian language usage; therefore, the 18 outputs classified as misinterpretations are deviations from these established norms due to various

⁴ Plotnikov, Timofei, 2024, "Replication Data for: Analyzing GPT-4 Misinterpretations of Russian Grammatical Constructions", <https://doi.org/10.18710/8CAPJM>, DataverseNO, V1

reasons, which are elaborated upon further. However, not all the 18 examples where GPT-4 struggles are due to misinterpretation of the fixed part of grammatical constructions. In some of the examples, there are other issues with the illustrative phrase, e.g., glitches, typos, and errors within the variable slot (a brief explanation of these misinterpretations is provided in *Section 4.2*.) While omitting the results that do not have issues with the fixed part of grammatical constructions, GPT-4 struggles with the correct interpretation of nine illustrative phrases across various language reference levels.

TABLE 1 - Incorrect interpretations of all examples versus incorrect interpretations of the examples with the fixed part of grammatical constructions assigned to the CEFR levels of language proficiency in RusCon

Language reference level	Number of grammatical constructions in RusCon	Overall misinterpretation of the examples	Misinterpretation of the examples with the fixed part of grammatical constructions
A1	86	0	0
A2	302	1	1
B1	654	5	1
B2	697	10	7
C1	428	1	0
C2	110	1	0
Percentage		0.8%	0.4%

(Source: Plotnikov, Timofei, 2024, "Replication Data for: Analyzing GPT-4 Misinterpretations of Russian Grammatical Constructions", <https://doi.org/10.18710/8CAPJM>, DataverseNO, V1)

While looking at the table, there were no issues with the interpretation of grammatical constructions within A1 level, just one inaccuracy within A2, one within B1, and seven within B2; there were additionally no issues within C1 and C2. Besides, the C2 level provides 110 construction examples, while B2 provides 697, which is nearly seven times more. Naturally, the larger dataset would also contain more misinterpretations in GPT's outputs, and that is what is observed between the B2 and C2 levels.

Since there were issues with the fixed part in nine grammatical constructions, five additional examples featuring the same grammatical construction were subsequently examined. However, only seven out of the nine grammatical constructions, where the issue occurred in the fixed part, included these five

additional examples. The initial assumption was that testing these extra examples could help draw more insightful conclusions regarding the misinterpretations of the nine grammatical constructions. Yet, these 35 additional examples are typically extensive and unambiguous, or they are a combination of two sentences, which leads to almost always correct and accurate interpretations due to the sufficient contextual information provided. Nevertheless, GPT-4 exhibits challenges in accurately interpreting just one additional example of the grammatical construction ID1652, which is discussed in more detail later

3.1. Misinterpretations of the fixed part

- (1) ID131 NP-Nom Cop zdorov VP-Inf
On zdorov vrat'
[he.NOM healthy lie.INF]
'He is good at lying'

GPT-4 states “the grammar and word choice are incorrect”⁴. However, GPT-4 identifies the meaning of the construction accurately but still insists that the word *zdorov* would not be used in the context of lying in standard Russian. Upon examination through Google Book Search, this example appears in eight attestations, including books and dictionaries; in all eight attestations, the construction is used in the context of lying. However, RusCon indicates that the construction is colloquial, signifying its rarity in formal written language. This results in insufficient training data with this grammatical construction in the context of lying, which is why GPT-4 struggles.

- (2) ID217 neznamo PronInt
Neznamo kak ja vernulsja domoj
[unknown how I return.PST.M homeward]
'Without knowing how I came back home'

GPT-4 identifies the usage of the grammatical construction with the adverb *neznamo* as an error, suggesting the substitution with *neizvestno* instead. According to *Bol'šoj tolkovyj slovar' russkogo jazyka* (Kuznecov, 2000: 623), *neznamo* is colloquial rather than formal, and thus, it is not commonly used in written discourse. Furthermore, data from the Russian Web 2011 corpus reveals that *neznamo* appears in 3,292 attestations, while *neizvestno* is found in 366,697 attestations. The significantly lower frequency of *neznamo* explains why GPT-4 recognizes it as an error, due to the scarcity of data involving this word, which is rare in written texts.

(3) ID587 VP po NP-Dat.Pl

Oni ljubili často ezdit' po znakomym
[they.NOM love.PST.PL often visit.INF by acquaintances.DAT]
'They often liked to visit their acquaintances'

Although GPT-4 interprets the meaning of the example correctly, it classifies the example as "somewhat ambiguous"⁴ due to the use of the preposition *po* followed by the word *znakomym*, which can function as both a plural noun and an adjective in varying contexts. Consequently, ambiguity arises, since GPT-4 interprets *znakomym* as an adjective and thus it anticipates an additional word after it.

(4) ID603 NP-Pl *odin drugoj*-Gen Adj-Cmp

Devočki odna drugoj strojnee
[girls.NOM one.F another.DAT slimmer.COMP]
'The girls are each slimmer than the other'

GPT-4 claims that the example is "grammatically incorrect and may cause confusion as to what exactly is being compared"⁴. However, additional examples in RusCon employing this grammatical construction incorporate either a dash or a comma before *odin drugoj*. Therefore, introducing a punctuation mark (a dash in this case in this particular example) could clarify the intended meaning by indicating that *devočki* is the subject of the sentence and the rest are related elements. Nevertheless, the absence of a dash is not a sufficient basis for GPT-4 to state the

grammatical structure entirely incorrect. Additionally, GPT-4 mentions that it is “difficult understand the intended meaning without additional context”⁴.

(5) ID1275 VP vvidu NP-Gen

On ne prišël vvidu bolezni
[he.NOM not come.PST.M due.to illness.GEN]
'He did not come due to illness'

GPT-4 suggests that the usage of *vvidu* is outdated and a more common way to use *iz-za* in this case, because *vvidu* is not used to indicate a cause or a reason in modern Russian and overall is considered too formal. However, according to *Slovar' trudnostej russkogo jazyka* (Rozental & Telenkova, 2005: 76-77) *vvidu* is used both to indicate a reason expected in the future, and to designate reasons that are timeless or related to the present or the past. Nevertheless, in the Russian Web 2011 corpus, *vvidu* can be found in 433,466 attestations, while *iz-za* appears in 3,987,006 attestations. The usage of *iz-za* is indeed more common, but *vvidu* is not outdated and still widely used in formal and more nuanced contexts. GPT-4's misinterpretation of this example stems from an overemphasis on the frequency of usage, which oversimplifies language nuances.

(6) ID1586 kogda XP, (a) kogda XP

Kogda vovremja pridët, kogda opozdaet
[when on.time come.3SG when late.arrive.3SG]
'Sometimes he/she/it arrives on time, sometimes he/she/it arrives late'

GPT-4 states that the structure of the example is incorrect. In the revised examples proposed by GPT-4, pronouns such as “someone” or “you” are incorporated, since grammatical patterns in English typically require a subject before a verb, which is not always necessary in Russian. Furthermore, GPT-4 removes the fixed part of the grammatical construction, thereby altering the intended meaning:

- *Kto-to pridët vovremja, a kto-to opozdaet* 'Someone will arrive on time, while someone else will be late'
- *Inogda prihodiš'vovremja, inogda opazdyvaeš'* 'Sometimes you arrive on time, sometimes you are late'

While attempting to convey the intended meaning, GPT-4 appears to impose an English syntactic structure, since the absence of a subject makes interpretation more confusing and complicated for GPT-4. In addition, GPT-4 once again indicates that it requires more context for a more accurate interpretation.

(7) ID1652 *ne segodnja zavtra* VP

Ne segodnja zavtra my vsë uznaem
[not today tomorrow we everything learn.1PL]
'One of the next few days we will found out everything'

GPT-4 claims that the example is incorrect. Upon examination, it appears that a more prevalent usage of this construction involves inserting a dash between *segodnja* and *zavtra* (i.e., *ne segodnja-zavtra*). For instance, the Russian Web 2011 corpus provides only 883 attestations for *ne segodnja zavtra* and 3,157 attestations for *ne segodnja-zavtra* when the dash is employed. However, RusCon states that the spelling employing a dash does not correspond to the modern spelling norm, despite the higher frequency in the corpus. As a correct example, GPT-4 recommends separating *ne segodnja* and *zavtra* with a comma (i.e., *ne segodnja, zavtra*), which can be accurate in some contexts but not in this specific case, as *ne segodnja zavtra* is an idiomatic phrase with its own distinct pragmatic meaning. According to RusCon, *ne segodnja zavtra* means that the action will take place in the next few days without an exact indication of a specific day, while *ne segodnja, zavtra* literally means 'not today, but tomorrow', which clearly indicates that the action will take place tomorrow.

In one of the five additional examples featuring this grammatical construction in RusCon (*k s'emkam fil'ma Petrov nameren byl pristupit' ne segodnja zavtra*), GPT-4 also suggests employing a dash or a comma in *ne segodnja zavtra*, which indicates that GPT-4's interpretation is influenced by frequency.

(8) ID1762 čto-to nezametno, čto Cl

Čto-to nezametno, čto ty toropiš'sja

[something not.noticeable that you.NOM hurry.2SG]

'It is not noticeable that you're in a hurry'

GPT-4 suggests that the current example is nonsensical. One of the revised examples proposed by GPT-4 is *kažetsja, ty nezametno toropiš'sja*, which removes the fixed part of the grammatical construction; the literal translation 'it seems you are hurrying unnoticeably' does not convey the intended meaning, indicating that GPT-4 has misinterpreted it. However, GPT-4 states that the example is not used in formal written discourse, and RusCon also indicates that the current construction's usage is colloquial. Additionally, the Russian National Corpus provides just three attestations with the fixed part of this grammatical construction, i.e., *čto-to nezametno, čto*, while ruTenTen 2011 does not provide any at all. Therefore, GPT-4 lacks sufficient data to more accurately interpret this grammatical construction.

(9) ID2272 kakoj/kakoe (tam) Adv/Adj!

Kakoe tam bystro!

[what.NOM there fast]

'Far from fast!'

According to RusCon, the grammatical construction in this example is colloquial and expresses disagreement. However, GPT-4 completely misinterprets it, claiming that this example is an attempt to express that something is fast, which is incorrect. The construction with the same adjective appears in 18 attestations across Google Books, which indicates insufficient frequency of usage. Besides, GPT-4 states that a better response is not possible without additional context or details.

3.2. Misinterpretations of other results

GPT-4 does not seem to encounter challenges in properly interpreting the fixed parts of the following grammatical constructions, as it predominantly struggles with the variable slots:

(10) ID280 NP-Nom *prinjat'* NP-Acc *za* NP-Acc

On prinjal togo mužčinu za svoego otca
[he.NOM took.PST that.ACC man.ACC for own.GEN father.ACC]
He mistook that man for his father

(11) ID1562 VP *vo ves' opor/vo vsju pryt'/vo ves' duh*

On bežal vo ves' opor
[he.NOM ran.PST in all strength]
'He ran at full speed'

(12) ID2040 ot NumCrd-Gen *otnjat'* NumCrd-Acc *ravno*/VP NumCrd-Nom/NumCrd-Dat

Ot vos'mi otnjat' tri budet pjat'
[from eight.GEN subtract.INF three be.FUT five]
'Subtracting three from eight equals five'

In all three cases, GPT-4 identifies the sentence structure as a mistake, yet it consistently generates identical outputs. This appears to be a glitch in GPT-4.

(13) ID356 *čem by* VP-Inf, VP (*by*)

Čem by učit'sja, on guljaet
[what would learn.INF he.NOM walk.3SG]
'Instead of studying, he is walking around'

In this example, GPT-4 actually encounters difficulties with the grammatical construction and claims that the example is incorrect because it is not standard

Russian. However, according to RusCon, the construction is classified as obsolete, indicating that it is no longer prevalent in contemporary standard Russian. Therefore, GPT-4's response can be considered partially correct, as it reflects contemporary Russian language usage.

(14) ID577 NP-Nom Cop NP-Nom VP-Inf

On master gotovit'

[he.NOM master cook.INF]

'He is a master at cooking'

GPT-4 states that this example is not correct, which is not the case. RusCon characterizes this grammatical construction as colloquial, which again reassumes infrequent occurrence in the training data. As a result, this particular example lacks attestations in both the Russian Web 2011 corpus and the Russian National Corpus, but there were ten attestations in Google Books, confirming the scarcity of data available online for GPT-4 to more accurately interpret this example.

(15) ID609 PronPers-Dat NP-Dat Cop Pred

Mne ušam holodno

[to.me ears.DAT cold]

'My ears are cold'

GPT-4 claims that this example is grammatically incorrect and subsequently suggests an alternative, *mne holodno uši*, which is also incorrect. Despite the grammatical accuracy of the original example, it lacks attestations in corpora and Google Books, indicating its low frequency. Considering the infrequent occurrence of the example, GPT-4 proposes an incorrect version as an alternative, which appears to be a phenomenon of AI *hallucinations*⁵.

⁵ *AI Hallucinations* are false responses generated by an LLM that may seem correct but are actually nonsensical at all (Rawte et al., 2023).

(16) ID808 VP v rezultate NP-Gen

Neskol'ko čelovek postradalo v rezultate vzryva
[several persons suffer.PST.PL in result.GEN explosion.GEN]
'Several people were injured as a result of the explosion'

GPT-4 states that *neskol'ko čelovek* is plural and that the verb should correspond in number with the subject. However, in Russian, verb-subject agreement in number can vary depending on the context and the speakers' intentions (Neset & Janda, 2023).

(17) ID1162 Adv-Cmp vsego VP

Èti busy podhodjat k plat'ju lučše vsego
[these beads suit.3PL to dress better of.all]
'These beads match the dress best'

(18) ID2084 bez NumCrd-Gen (minut) NumCrd-Nom

Prihodi k bez pjatnadcati desjat'
[come.IMP to without fifteen ten]
'Come by fifteen to ten'

Even though GPT-4 misinterprets these examples, it does not claim that these examples are entirely incorrect. It suggests that the incorporation of the preposition *k* might be slightly ungrammatical, as this preposition is not typically required after the verbs in these examples.

4. Discussion

The analysis indicates that GPT-4 demonstrates a high level of proficiency in handling Russian grammatical constructions. Among 2,227 constructions, GPT-4 struggled with 18 constructions, representing less than 1% of the entire dataset (0.4% referring to the misinterpretation of the fixed part of grammatical

constructions). Consequently, the response to the first question in the introduction is that GPT-4 demonstrates effective performance.

The additional inquiry concerned factors causing GPT-4 misinterpretations. Overall, GPT-4 tends to struggle with grammatical constructions when there is insufficient and infrequent data available (8 out of 18 examples), which evidences the frequency bias mentioned in the studies by Dudy & Bedrick (2020) and Brown *et al.* (2020). For other misinterpretations (4 out of 18), GPT-4 required more context necessary for accurate input interpretation, since GPTs can struggle to perform effectively due to contextual deficits (Gong *et al.*, 2019; Liu *et al.*, 2021; Yenduri *et al.*, 2024). Predominantly, the linguistic structures that challenge GPT-4 are those that are colloquial and infrequently represented in formal written discourse, thereby contributing to a scarcity of documented data available on the Internet.

Further, GPT-4 struggles with certain grammar patterns. It states that the grammar is incorrect in examples (1), (5), (16) and produces “correct” suggestions based on the frequency of the dataset rather than on grammatical accuracy. In example (6), GPT-4 appears to incorporate English syntactic structure while attempting to propose revised examples in Russian. GPT-4 also alters the original lexicon of the example (6), removing the fixed part of the grammatical construction; the same phenomenon is observed in example (8). This aligns with the findings of Papadimitriou *et al.* (2022) and Wendler *et al.* (2024) regarding the influence of the English language on the final output. In another example (7), GPT-4 inserts unnecessary punctuation marks in the output, confirming that LLMs can introduce irrelevant linguistic elements (Hendy *et al.*, 2023).

Except for the incorrect punctuation insertion, the example (7) has an idiomatic meaning, which GPT-4 cannot identify correctly. In example (9), GPT-4 also interprets the meaning literally rather than figuratively. These findings show that GPTs can exhibit limitations to interpret idiomatic expressions and non-literal language (Ray, 2023; Yenduri *et al.*, 2024). Furthermore, these examples demonstrate that LLMs encounter difficulties in representing cultural values in languages other than English (Johnson *et al.*, 2022; Jackson, 2023; Liu, 2024), considering that idiomatic expressions are not merely linguistics items but also carries of cultural significance.

In example (15), GPT-4 claims that the example is incorrect and attempts to provide an alternative, which is neither sensible nor standard Russian. This indicates that GPTs struggle to provide an adequate result when the initial input contains infrequent and rare words (Brown *et al.*, 2020; Dudy & Bedrick, 2020), potentially leading to AI hallucinations (Rawte *et al.*, 2023).

Some GPT-4 outputs with errors were not entirely wrong. GPT-4 claims that example (13) is not correct, but the grammatical construction is indeed archaic and not used in modern Russian anymore. Examples (17) and (18) were considered slightly incorrect due to a preposition, which is rather misleading as these examples are grammatically correct but are less frequently used in the Russian language. Therefore, GPT-4 suggests avoiding the preposition in those examples prioritizing word frequency over actual accuracy (Dudy & Bedrick, 2020).

Although GPT-4 demonstrates proficient performance across the majority of the dataset, some challenges remain, indicating that AI chatbots are not yet capable of completely substituting for human expertise. Researchers, teachers, and learners, who are non-native speakers, cannot fully rely on LLMs, since biased and inaccurate outputs can convey wrong linguistic ideas and might seem correct. To avoid such issues, expertise from human professionals is still essential. However, one solution for achieving better outputs is to provide more context. My experiment particularly demonstrates this in the testing of five additional examples of grammatical constructions where issues occurred in the fixed part. The results were mostly appropriate due to the sufficient context provided.

Regarding LLM development, it might be hard to fix the frequency bias at the present stage, but one of the solutions is to have more balanced training data. When the data is predominantly in English and data in Russian, which is one of the most widely spoken languages, represents only 0.2% in the main text corpus implemented in GPTs, the situation is less than ideal and needs attention from LLM developers. Although GPT-4 interprets Russian grammatical constructions quite well, its performance might be significantly worse with less commonly spoken languages that have structures vastly different from English.

5. Limitations and future directions

In the present study, the experiment was conducted using the AI chatbot ChatUiT, which functions similarly to ChatGPT but is not available to the general public. However, the results are generalizable because ChatUiT implements the foundational model GPT-4, the same one that is being employed in ChatGPT at the time the experiment was conducted.

Additionally, a zero-shot experiment was conducted in this study. While examples (10), (11), and (12) appear to be products of AI hallucinations, GPTs' responses may potentially vary if semantically/syntactically similar examples are provided in advance before conducting an analogous experiment. Thus, the "problematic" dataset identified in this investigation can also be tested with such examples.

There were also a few limitations when comparing GPT-4 interpretations with annotations from RusCon. Some entries had minor errors, occasional typos, or lacked sufficient information, since the RusCon project was completed manually rather than computationally. Furthermore, the dataset in RusCon is not large; additional constructions could potentially reveal more inaccuracies in areas other than those already mentioned.

Overall, GPT models from OpenAI are not the only large language models available on the market. There are other prominent ones, such as Gemini from Google and Claude from Anthropic. The methods of data training and the settings differ among these LLMs. Therefore, it could be interesting to see and compare how various LLMs perform on the same task.

References

- Bast, R., Endresen, A., Janda, L. A., Lund, M., Lyashevskaya, O., Mordashova, D., Nettet, T., Rakhilina, E., Tyers, F. M., & Zhukova, V. (2021). *The Russian Constructicon. An Electronic Database of the Russian Grammatical Constructions*. <https://constructicon.github.io/russian/>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C.,

- McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *ArXiv abs/2005.14165*. Web.
- Diessel, H. (2019). *The Grammar Network: How Linguistic Structure Is Shaped by Language Use*. Cambridge University Press.
- Diessel, H. (2023). *The Constructicon: Taxonomies and Networks*. Cambridge University Press.
- Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M., & Gardner, M. (2021). Documenting large webtext corpora: A case study on the colossal clean crawled corpus. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 1286-1305). <https://doi.org/10.18653/v1/2021.emnlp-main.98>
- Dudy, S., & Bedrick, S. (2020). Are some words worth more than others? *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 131-142). <https://doi.org/10.18653/v1/2020.eval4nlp-1.13>
- Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, 64(3), 501-538.
- Gemini Team. Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.-B., Yu, J., Soricut, R., et al. (2023). Gemini: A family of highly capable multimodal models. *ArXiv preprint arXiv:2312.11805*.
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.
- Goldberg, A., & Suttle, L. (2010). Construction grammar. *WIREs Cognitive Science*, 1(4), 468-477. <https://doi.org/10.1002/wcs.22>
- Gong, X.-R., Jin, J.-X., & Zhang, T. (2019). Sentiment analysis using autoregressive language modeling and broad learning system. *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1130-1134).
- Google Books. (n.d.). Retrieved July 15, 2024, from <https://books.google.com/>
- Hendy, A., Abdelrehim, M. G., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M. A., & Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. *ArXiv abs/2302.09210*.
- Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh University Press.
- Jackson, S., Beekhuizen, B., Zhao, Z., Zhao, Y. C., & McEwen, R. N. (2023). LLMs and linguistic competency: An exploration of GPT-4 and a non-hegemonic English variety. *Newhouse Impact Journal*, 1(1), 21-23. <http://doi.org/10.14305/jn.29960819.2024.1.1.04>
- Janda, L. A., Endresen, A., Zhukova, V., Mordashova, D., & Rakhilina, E. (2020). How to build a constructicon in five years: The Russian example. *Belgian Journal of Linguistics*, 34, 162-175. <https://doi.org/10.1075/bjl.00043.jan>
- Janda, L. A., Lyashevskaya, O., Nessel, T., Rakhilina, E., & Tyers, F. M. (2018). Chapter 6. A constructicon for Russian: Filling in the gaps. In B. Lyngfelt, L. Borin, K. Ohara, & T. T. Torrent (Eds.), *Constructicography: Constructicon development across languages* (pp. 165-181). John Benjamins Publishing Co. <https://doi.org/10.1075/cal.22.06jan>
- Janda, L. A., Zhukova, V., & Endresen, A. (2024). Typology of reduplication in Russian: Constructions within and beyond a single clause. In M. Kopotev & K.

- Kwon (Eds.), *Constructions with lexical repetitions in East Slavic* (pp. 71-96). De Gruyter Mouton. <https://doi.org/10.1515/9783111165806-003>
- Johnson, R. L., Pistilli, G., Menéndez-González, N., Dias Duran, L. D., Panai, E., Kalpokiene, J., & Bertulfo, D. J. (2022). The ghost in the machine has an American accent: Value conflict in GPT-3. *arXiv preprint arXiv:2203.07785*.
- Kuznecov, S. A. (Ed.). (2000). *Bol'shoj tolkovyj slovar' russkogo jazyka*. Bukinist.
- Lai, V. D., Ngo, N. T., Pourn Ben Veyseh, A., Man, H., Derroncourt, F., Bui, T., & Nguyen, T. H. (2023). Chatgpt beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. *arXiv preprint arXiv:2304.05613*.
- Liu, C. (2024). The investigation of application related to ChatGPT in foreign language learning. *Applied and Computational Engineering*, 35, 110-115. <https://doi.org/10.54254/2755-2721/35/20230376>
- Liu, J., Shen, D., Zhang, Y., Dolan, B., Carin, L., & Chen, W. (2021). What makes good in-context examples for GPT-3? Workshop on Knowledge Extraction and Integration for Deep Learning Architectures; Deep Learning Inside Out. *arXiv preprint arXiv:2101.06804*.
- Lyngfelt, B., Borin, L., Ohara, K., & Torrent, T. T. (Eds.). (2018). *Constructicography: Constructicon development across languages*. John Benjamins. <https://doi.org/10.1075/cal.22>
- National Corpus of the Russian Language. (2024). Retrieved May 12, 2024, from <https://ruscorpora.ru/en/>
- Nesset, T., & Janda, L. A. (2023). A network of allostructions: Quantified subject constructions in Russian. *Cognitive Linguistics*, 34(1), 67-97. <https://doi.org/10.1515/cog-2021-0117>
- Pannatier, A., Courdier, E., & Fleuret, F. (2024). GPTs: A new approach to autoregressive models. *arXiv preprint arXiv:2404.09562*.
- Papadimitriou, I., Lopez, K., & Jurafsky, D. (2022). Multilingual BERT has an accent: Evaluating English influences on fluency in multilingual models. *arXiv abs/2210.05619*.
- Rakhilina, E., Zhukova, V., Demidova, D., Kudrjavceva, P., Rozovskaja, G., Endresen, A., & Janda, L. A. (2022). Frazеologija v rakurse Russkogo Konstruktikona [Phraseology in the light of the Russian Constructicon]. *Bulletin of the Russian Academy of Sciences: Studies in Literature and Language*, 2(32), 13-44. <https://doi.org/10.31912/pvrli-2022.2.2>
- Rawte, V., Chakraborty, S., Pathak, A., Sarkar, A., Tonmoy, S. M., Chadha, A., Sheth, A. P., & Das, A. (2023). The troubling emergence of hallucination in large language models - An extensive definition, quantification, and prescriptive remediations. *arXiv preprint arXiv:2310.04988*.
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Rozental', D.Ě., & Telenkova, M.A. (2005). *Slovar' trudnostej russkogo jazyka*. Ajrispress.
- Sketch Engine - Russian Web 2011 corpus (ruTenTen 2011). (2024). Retrieved May 12, 2024, from <https://app.sketchengine.eu>

- Tan, Y., Min, D., Li, Y., Li, W., Hu, N., Chen, Y., & Qi, G. (2023). Can ChatGPT replace traditional KBQA models? An in-depth analysis of the question answering performance of the GPT LLM family. *International Semantic Web Conference* (pp. 348-367). Cham: Springer Nature Switzerland.
- Wendler, C., Veselovsky, V., Monea, G., & West, R. (2024). Do llamas work in English? On the latent language of multilingual transformers. *arXiv abs/2402.10588*, 1-29. <https://arxiv.org/abs/2402.10588>
- Yenduri, G., Murugan, R., Govardanan, C., Supriya, Y., Srivastava, G., Reddy, P., Raj, D., Jhaveri, R., B, P., Wang, W., Vasilakos, A., & Gadekallu, T. (2024). GPT (Generative Pre-Trained Transformer) - A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 12, 54608-54649. <https://doi.org/10.1109/ACCESS.2024.3389497>
- Zhu, Q., & Luo, J. (2022). Generative pre-trained transformer for design concept generation: An exploration. *Proceedings of the Design Society*, 2, 1825-1834. <https://doi.org/10.1017/pds.2022.185>

Recensões

Teresa Brocardo In memoriam (1959-2023)

Isabel Margarida Duarte

Faculdade de Letras da Universidade do Porto (Portugal)

Centro de Linguística da Universidade do Porto (Portugal)

iduarte@letras.up.pt

"...Sedia la fremosa..." Textos de Teresa Brocardo. Coutinho, A., Correia, C. N. & Caetano, M. do C. (2024). Lisboa, Edições Colibri, 146 pp. ISBN 978-989-566-419-1.
<https://doi.org/10.34619/qyxj-kwdn>

Teresa Brocardo partiu demasiado cedo. Professora Catedrática do Departamento de Linguística da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa e investigadora do respetivo Centro de Linguística (CLUNL), do Grupo Gramática & Texto, na área da história da língua portuguesa e da linguística histórica, aliava conhecimentos profundos no seu campo de estudos com uma grande capacidade de problematização de questões teóricas complexas e talvez por isso fascinantes.

Além de dezenas de capítulos de livros e de artigos em revistas, deixou-nos dois livros: um, sobre o verbo, estudado de diferentes perspetivas linguísticas, em diferentes sincronias e tendo em conta a variedade portuguesa e a brasileira do português, organizado com Alvanira Barros, *Estudos sobre o verbo em português* editado pela Ideia, em 2016; o outro, a solo, intitulado *Tópicos de História da Língua Portuguesa*, publicado pela Colibri, em 2014.

Era, também, filóloga, tendo-lhe nós ficado a dever edições de alguns textos medievais, devidamente enquadradas por estudos linguísticos que os iluminam: *Crónica do Conde D. Pedro de Meneses de Gomes Eanes de Zurara. Edição e Estudo*, editada, em 1997, pela Fundação Calouste Gulbenkian e *Livro de Linhagens do Conde D. Pedro. Edição do fragmento manuscrito da Biblioteca da Ajuda (século XIV)*, publicado pela Imprensa Nacional - Casa da Moeda, em 2006.

Teresa Brocardo juntava a muitos conhecimentos e rigor outras qualidades que são inestimáveis: sabia muito, mas sem ponta de arrogância e era de uma grande gentileza e disponibilidade.

O livro *...Sedia la fremosa... Textos de Teresa Brocardo*, de que passaremos a fazer uma breve recensão, foi organizado por três colegas e amigas do mesmo grupo de investigação, Antónia Coutinho, Clara Nunes Correia e Maria do Céu Caetano, e editado pelas Edições Colibri em 2024, aquando de uma justa homenagem que lhe foi prestada, em 27 e 28 de junho de 2024, sob a forma de um *workshop* promovido pelo grupo Gramática & Texto, do Centro de Investigação de que fez parte. Nele, as organizadoras dão conta dessa ausência de vaidade, chamando a atenção para o facto de Teresa Brocardo se referir às suas intervenções nos *Workshops em Gramática e Texto* como “propostas de estudo” e “nótulas”, quando, na verdade, os seus contributos são de enorme importância para a área da história da língua e da linguística em geral.

O volume de que nos ocuparemos agora, também em jeito de homenagem, é composto por uma comovida Apresentação (“Estes textos - uma questão de *ser e estar* (no tempo)”, p. 7) da autoria das organizadoras, seguida de um interessante Texto de Abertura (“Grandes trezentos anos esteve assí, ou mais, cuidando que non estevéra senon pouco”, p. 11) da jornalista Sara Figueiredo Costa, orientanda de Teresa Brocardo. Interessante por nos mostrar, de forma heterodoxa, é verdade, como um professor pode fazer nascer, nos seus estudantes, a curiosidade intelectual e o entusiasmo relativamente a certos tópicos de investigação e interessante, ainda, pelo retrato humano que traça de Teresa Brocardo. Segue-se uma antologia de dez textos (“*“Sedia la fremosa...” Textos de Teresa Brocardo*”), pp. 23-115) originários dos *Workshops em Gramática e Texto* e uma Bibliografia da autora (p. 115). Na última parte do livro, intitulada “Leituras”, juntam-se textos de cada uma das organizadoras, sobre Teresa Brocardo: “Ler Teresa Brocardo”, de Clara Nunes Correia (p. 127), “Quantas margens tem um texto? - divagações e interpelações” (p. 133), de Antónia Coutinho, e “Formas e construções em diacronia” (p. 141), de Maria do Céu Caetano. Cada uma delas se centra nos textos da homenageada que melhor dialogam com os seus próprios interesses de

investigação e será assim que os vamos apresentar: também em diálogo com os da homenageada.

A preocupação do grupo em cruzar pontos de vista por vezes discordantes, procurando que umas investigadoras instigassem as outras a ir mais longe na pesquisa, fica patente em alguns dos textos de Teresa Brocardo que agora se publicam, por exemplo, em “Nótulas históricas - uma (re)leitura de Campos (2000)” (p. 43), em que parte de um texto de Henriqueta Costa Campos sobre o Pretérito Perfeito Composto, para um interessantíssimo artigo sobre a especificidade desse tempo em português, por contraponto com outras línguas românicas, servindo-se de uma visão diacrónica e dos valores de construções com *haver* e *ter*. A autora volta a este par num outro estudo, “*Haver e ter - Balanços e perspetivas*” (p. 101), em que se ocupa destes verbos, na história do português, “em construções que desencadeiam leituras caracterizáveis como epistémicas, no sentido em que marcam alguma expressão de validação” (2024, p. 103) e são usados, embora não muito frequentemente, ainda hoje. A relação entre estes valores epistémicos e outros funcionamentos de *haver* e *ter* seria também um tópico a explorar no futuro, mas a autora adianta já algumas pistas: os verbos ocorrem em estruturas transitivas predicativas “em que coocorrem formas / expressões lexicais [...] que favorecem a inferência do valor epistémico da construção, e que corresponderá a uma validação “forte” do asserido”. (2024, p. 105). Os exemplos utilizados confirmam, sem dúvida, esta hipótese de explicação.

O diálogo com os ensinamentos de Henriqueta Costa Campos volta a acontecer no capítulo final, “Nas «Margens» do «texto» - normas sobre variantes em formas de *dever* e *poder*” (p. 107). No estudo revisitado de Campos (2000), esta analisa o pretérito perfeito e o mais-que-perfeito, simples e composto, usando passagens da *Crónica de D. João I* de Fernão Lopes, e defende que o trabalho sobre história da língua pode ter relevância para o estudo dos valores contemporâneos dessas formas verbais. Teresa Brocardo discute, por seu lado, o valor dos verbos modais (deônticos ou epistémicos), referidos a partir de variantes, quer dizer, de dois dos “testemunhos” de outra crónica, a de Zurara: *Crónica do Conde D. Pedro de Meneses*, mostrando a utilidade e a validade deste tipo de confronto de versões diversas.

Clara Nunes Correia, no seu texto da parte final do livro, já referido, “lê” estes textos de Teresa Brocardo sobre tempos verbais (por exemplo, sobre valores temporais e modais do futuro do pretérito/condicional), convocando também os textos de Henriqueta Costa Campos referidos por Brocardo. A propósito dos verbos modais *poder* e *dever*, e do seu valor inicialmente deôntico, Clara Correia escreve: “Mais do que uma descrição de ocorrências, Brocardo ensina-nos a ver a relevância das fontes para a compreensão das formas e das construções em análise.” (2024, p. 129). Refere-se ainda ao texto de Teresa Brocardo “Representações do / no tempo - nótula sobre *passados*” (p. 85). Nele, a autora descreve a diacronia de dois paradigmas verbais do passado: o futuro do passado/ condicional e o pretérito mais-que-perfeito simples, com valores quer temporais quer modais, para o que convoca não só a noção de “competição”, mas também os conceitos de inovação e conservação ou persistência, e o de obsolescência. Uma das lições que importa ressaltar, deste texto, é que “[...] mesmo uma mera descrição da mudança linguística não pode nunca centrar-se numa dada unidade (forma, construção, paradigma...) sem considerar outra ou outras que com ela competem numa dada sincronia.” (2024, p. 90).

Como se disse, a investigação de Teresa Brocardo, centrada na história da língua portuguesa, traz um contributo inestimável para a Linguística em geral, nomeadamente porque questiona e aprofunda um conjunto de conceitos teóricos de interesse para todas as áreas, como a “explicação” em diacronia, a partir da qual defende o cruzamento de contributos de abordagens sincrónicas e diacrónicas para uma compreensão mais capaz dos fenómenos linguísticos (cf. o texto “...Porque *foi* assim! (Notas sobre a explicação em diacronia)” (p. 33)).

Um outro texto do pequeno livro a que nos vimos referindo atesta esta vocação de problematizar grandes questões teóricas: “Polissemia e mudança linguística” (p. 23), em que a autora sublinha a necessidade de uma fase de polissemia como condição obrigatória para que a mudança semântica ocorra. Esta problematização é feita com recurso ao caso concreto de “*haver*” em português, como aliás ocorre sempre nestes textos, em que Teresa Brocardo trata de noções teóricas, mas a partir de exemplos da história da língua portuguesa. É o que acontece, também, em “Forma, construção e significado - nótula sobre

gramaticalização e analogia” (p. 53). Nesse texto, fica visível, mais uma vez, o contraste entre a profundidade dos seus conhecimentos e do nível de problematização e a modéstia com que a autora encarava o seu próprio contributo. Leia-se a seguinte passagem, como testemunho disto mesmo:

A apresentação propõe como hipótese de trabalho (muito) preliminar, uma abordagem do processo de gramaticalização de construções do português com *haver / ter + PP*, insistindo na ideia, já formulada em trabalhos anteriores (p.ex., Brocardo 2009), de uma motivação analógica do referido processo, desencadeado pela co-ocorrência de construções formalmente paralelas, já plenamente gramaticalizadas na expressão de valores têmporo-aspectuais (perífrases com formas derivadas de ESSE + PP). (2024, p. 54).¹

A partir, pois, de exemplos específicos da história da língua, a autora discute a seguinte questão: “sobre que unidade, afinal - forma ou construção - incidem os processos de gramaticalização?” (2024, p. 59). A esta seguem-se mais três perguntas bem feitas, numa secção final do texto que Teresa Brocardo intitulou, com humor, “Não há conclusões”. Mas, em todo o caso, com bom senso, vai propondo que se equilibrem ‘macro’ e ‘micro’ análises, ou seja, análises de períodos longos, por vezes simplificadoras, com análises pormenorizadas, necessariamente com menos dados.

Maria do Céu Caetano centra a sua reflexão sobre questões dos estudos diacrónicos e dialoga sobretudo com três textos de Teresa Brocardo: aquele em que é tratada a noção de polissemia e mudança linguística; o texto em que Brocardo equaciona o valor da “explicação” para compreender essa mudança, e aquele em que, depois de discorrer sobre gramaticalização e analogia, a partir de dados da diacronia do português, a autora argumenta de modo a que se conclua que “nem sempre é possível isolar os mecanismos e os resultados de processos de gramaticalização motivados analogicamente dos de processos de “pure grammaticalization” [...]” (2024, p. 146).

Este volume dá conta do cruzamento de perspetivas dentro do grupo de investigação. É disso que se trata em “O texto no tempo em que não havia gramáticas” (p. 63). Partindo da ideia de que o texto tem interesse enquanto objeto

¹ A ortografia dos textos publicados antes da entrada em vigor do Acordo Ortográfico não foi atualizada.

tanto para a Linguística de Texto quanto para a linguística histórica, já que é nos textos que se colhem exemplos de dados para trabalhar em diacronia, Teresa Brocardo mostra como a noção de género, vinda da primeira, poderia ser útil para o trabalho da segunda. Este seu artigo, em que distingue *texto* e *testemunho*, vai ser, aliás, o ponto de partida para a reflexão entusiástica que Antónia Coutinho fará sobre a obra de Teresa Brocardo, na parte final do volume. Antónia Coutinho revisita sobretudo artigos em que Teresa Brocardo se refere à noção de texto e à sua importância para os estudos diacrónicos, problematizando essa noção no que se refere aos *testemunhos* a que a história da língua e a linguística histórica não conferem o estatuto de *texto*, posição que Antónia Coutinho de certo modo contesta. Teresa Brocardo sublinha a importância de se olhar para os textos antes de se olhar para os *corpora*, e esse movimento metodológico já tem total aprovação da sua colega do grupo de investigação.

No texto seguinte, "*Sedia la fremosa... Uma proposta de estudo diacrónico de ser (< sedere e esse) e estar em português*" (p. 75), Teresa Brocardo pretende "delinear uma proposta de estudo sobre a diacronia de *ser* e *estar* em português" (2024, p. 76). Sublinha, como habitualmente, o carácter de mera proposta do seu trabalho, embora a complexidade da sua reflexão merecesse outra designação mais justa. Na verdade, sugere todo um percurso de análise fina que permite conhecer melhor os valores subtis por que os usos dos referidos verbos passaram.

Em "Comente o seguinte *texto-fonte*"² (p. 93), Teresa Brocardo propõe "algumas notas" sobre esta atividade, salientando aspetos centrais sobre "o texto no tempo em que não havia gramáticas", ou seja, debruçando-se sobre questões metodológicas decorrentes "da exploração linguística de testemunhos escritos de fases passadas da língua para as quais não existem fontes 'secundárias'" (2024, p. 93). É um estudo mais de carácter filológico, que recorre a testemunhos do *Testamento de Afonso II* (1214) e usa, como breve exemplo, o possessivo precedido de artigo, que já aparece nesse documento, questionando generalizações feitas por Rinke (2010) sobre o mesmo tópico. Aconselha, a nosso ver convincentemente, que

² É nítida a ressonância do romance *Comente o seguinte texto*, de Eduarda Dionísio, editado em Lisboa, pela Plátano Editora, em 1972. O título deu nome a um dos *Workshops* do grupo Gramática & Texto, em 2016.

se olhe para o texto antes de se olhar para o *corpus*, ou, dito de outro modo, que o uso de *corpora* quantitativamente pertinentes não faça esquecer a necessidade de “começar pelo texto, ou antes ainda, quando possível, pelos seus vários testemunhos.” (2024, p. 99).

Feita esta breve recensão ao livro de homenagem a Teresa Brocardo, não podemos esquecer, além do valioso contributo da sua investigação, sublinhado pelas organizadoras da obra em apreço, uma outra vertente da homenageada: a sua participação generosa na governação universitária. Quando nos deixou, era vice-presidente do Conselho Científico e subdiretora para a Gestão Curricular e Estudantes, mas ocupou vários outros cargos na sua instituição de pertença. Foi presidente do Conselho Pedagógico (2002-2005) e coordenadora do Departamento de Linguística. Falta acrescentar, em jeito de nota pessoal, a imagem que de Teresa Brocardo guardo, do convívio que com ela tive, enquanto membro do mesmo painel de avaliação de Bolsas de Doutoramento da Fundação para a Ciência e a Tecnologia ou de vários júris universitários: a de uma pessoa serena, com muito bom senso, sentido de humor e uma enorme disponibilidade para os outros. Fica-nos essa imagem e os textos publicados, fruto de “honesto estudo, com longa experiência misturado” e engenho, textos que será sempre estimulante visitar.

Referências

- Campos, M.H.C. (2000). Sur les formes composées du prétérit em portugais. In A. Englebert et al. (Eds). *Actes du XXII^e Congrès International de Linguistique et Philologie Romanes* (vol. II, pp. 57-63).
- Rinke, E. (2010). A combinação de artigo definido e pronome possessivo na história do português. *Estudos de Lingüística Galega*, 2, 121-139.

Notas

Oswald Ducrot (27/11/1930 - 08/06/2024)

Isabel Margarida Duarte

Faculdade de Letras da Universidade do Porto (Portugal)

Centro de Linguística da Universidade do Porto (Portugal)

iduarte@letras.up.pt

Oswald Ducrot partiu aos 93 anos, em 8 de junho de 2024. Deixa-nos um interessantíssimo legado no campo da Linguística, mais especificamente nos estudos da teoria da argumentação. Mas não só. Com Tzvetan Todorov, organizou o *Dictionnaire encyclopédique des sciences du langage*, editado pela Seuil em 1972, uma obra abrangente e de grande utilidade. Como se escreve na apresentação, "Ce dictionnaire ne se limite pas à la linguistique *stricto sensu*; y figurent aussi les concepts fondateurs, comme celui de signe, et, symétriquement, on y a pris en considération les productions de la langue (d'où la place accordée à la poétique). Quatre grandes parties: les écoles (depuis le xv^e siècle jusqu'à Chomsky), les domaines (y compris psycho et socio-linguistique), les concepts méthodologiques (du plus fondamental - le signe - au plus dérivé - les genres littéraires), les concepts descriptifs (du plus simple - les unités non significatives - aux plus complexes - du langage et de l'action)." Mas seria uma outra parceria, desta vez com Jean-Claude Anscombe, que iria ficar para a história dos estudos de argumentação na linguagem. Em *L'argumentation dans la langue* (Anscombe & Ducrot, 1976, 1983) os autores defendem que o principal objetivo da linguagem não é representar estados de coisas, mas sim argumentar. A estrutura da argumentação faz parte da própria língua. Daí decorre uma análise muito interessante e frutífera sobre elementos e estruturas da língua que orientam a argumentação num determinado sentido, como os conectores e articuladores discursivos, por exemplo. Segundo a teoria da argumentação, que não considera os contextos nem quaisquer fatores externos à língua, inserindo-se, portanto, numa tradição estruturalista, faz parte dos enunciados e da respetiva articulação uma vocação para a orientação argumentativa. A língua indica, através de mecanismos próprios, a direção que o locutor pretende sugerir ao respetivo destinatário. Por exemplo, perante o

enunciado “O livro é grande, mas interessante”, a continuação previsível irá no sentido de que vale a pena lê-lo. Mas o enunciado “O livro é interessante, mas grande” deixa prever uma continuação que não aconselha a sua leitura. Por outro lado, a própria seleção do léxico é argumentativa, no sentido em que o adjetivo “grande”, qualificando “livro”, não remete para uma qualidade totalmente objetiva, mas antes para uma avaliação, neste caso com coloração negativa. Ou seja, a palavra escolhida é, também, uma forma de orientar a argumentação.

Os usos e valores argumentativos de conectores como “mas” e articuladores como “décidément”, “Eh bien!”, “d’ailleurs”, bem como verbos de opinião tal como “je trouve que” foram estudados em *Les mots du discours*. Em *Les échelles argumentatives*, Ducrot (1980, 12) afirma: “Toute parole, au fond d’elle-même, est publicitaire. Elle n’est pas publicitaire seulement par le fait qu’elle véhicule certaines informations qui se trouvent autoriser certaines conclusions. Elle est publicitaire par le fait que sa valeur interne se confond avec la suite qu’elle réclame. Ce qu’elle veut dire, c’est ce qu’elle veut faire dire à l’autre. Ainsi nos énoncés se présentent, indépendamment même de leur aptitude à fonder un raisonnement, comme l’origine ou le relais d’un discours argumentatif.” O seu estudo sobre as diferenças argumentativas entre “peu” e “un peu” é um excelente exemplo de como um dado elemento da língua tem importância decisiva para guiar a interpretação do destinatário numa determinada direção argumentativa.

Uma outra área em que Ducrot nos deixou um legado muito interessante é a do subentendido, do pressuposto e do posto, das diferenças entre o dito e o comunicado. Ocupou-se da distinção entre o que dizemos e o que é pressuposto por esse dizer, isto é, da distância entre as informações explícitas e o que subentendemos, sugerimos ou pressupomos. Distinguiu, assim, o pressuposto, o posto e o subentendido. O pressuposto está, segundo Ducrot, inscrito no próprio enunciado. Se usarmos o exemplo clássico de Bertrand Russell “O atual rei de França é careca”, tal enunciado pressupõe que existe um rei de França atualmente e o posto diz-nos que ele é careca. Se eu negar ou interrogar o enunciado, o pressuposto não se altera (existe um rei de França), mas apenas o posto. Quanto ao subentendido, não decorre de nenhum elemento linguístico e podemos, portanto, sempre afirmar que não foi dito, mas apenas sugerido. A responsabilidade da sua

existência recai sobre o destinatário e não sobre o locutor. O autor explora, por exemplo em *Dire et ne pas dire*, que tem por subtítulo *Principes de sémantique linguistique*, os subentendidos e as pressuposições, lembrando que “[...] on a bien fréquemment besoin, à la fois de dire certaines choses, et de pouvoir faire comme si on ne les avait pas dites, de les dire, mais de façon telle qu’on puisse refuser la responsabilité de leur énonciation.” (Ducrot, 1980, 5). A sua teoria tem óbvios pontos de contacto com a das implicaturas de Grice (1975) e sinaliza uma área da Pragmática cujos desenvolvimentos teóricos não cessaram de crescer desde então e a que Levinson (1983) chamaria Pragmática de segundo grau. Se o autor fala de Semântica, esta tem, no entanto, um pendor claramente enunciativo e pragmático. Muitos dos tópicos trabalhados por Ducrot permitem justamente problematizar os limites entre Semântica e Pragmática.

Um terceiro conceito que o autor retomou e problematizou do ponto de vista da Linguística, com enorme proveito, foi o de polifonia. Não polifonia no sentido exato de Bakhtine (1929), uma polifonia que existiria sobretudo no romance, onde várias vozes, das personagens e do locutor contracenariam, mas inspirando-se, também, em Bally (1944), que já falara em diferentes vozes presentes na língua. O enunciado é marcado, como sabemos, por várias vozes, nomeadamente as que remetem para a deixis pessoal, isto é, as vozes dos intervenientes na enunciação. A presença dos interlocutores no enunciado é uma irrupção do discurso na língua. Explicitamente inspirado por Gérard Genette e pelas suas distinções célebres entre autor, narrador, personagem, locutor, Ducrot sublinha a existência de diferentes vozes no enunciado, além das marcas de 1.^a e 2.^a pessoas: a do autor empírico, a do locutor e a de vários possíveis enunciadores com diferentes pontos de vista. O locutor é o responsável pela enunciação e pode assumir a voz de um dado enunciador ou pode afastar-se dela, assim conseguindo diferentes efeitos, por exemplo, na negação, que pressupõe um enunciador capaz de produzir o enunciado afirmativo correspondente ao negativo: se o locutor diz “Agora não chove”, imagina-se que um qualquer enunciador poderia dizer ou pensar que “Agora chove”. Mas também a ironia é facilmente explicável pela noção de polifonia de Ducrot: o locutor afastar-se-ia de um enunciado cuja avaliação apreciativa pertence a um enunciador, diferente de si, do qual a sua apreciação se afasta.

Portanto, a teoria da polifonia, tal como a entende, tem uma orientação instrucional relativamente ao destinatário a quem o texto indica um caminho argumentativo. Esta teoria vai alcançar grande sucesso quando se cruza com a noção de ponto de vista, não só na literatura, mas, sobretudo, nos discursos quotidianos, atravessados por fortes dimensões argumentativas. O conceito de polifonia de Ducrot permitiu ainda avanços no que diz respeito ao relato do discurso e foi glosado por vários linguistas da área da Análise do Discurso.

Vale a pena considerar ainda a noção de *ethos* do autor (herdada de Aristóteles) que decorre da sua teoria da polifonia, uma vez que a associa ao locutor e não ao sujeito empírico. Segundo o autor, os modos de dizer do locutor caracterizam-no melhor do que aquilo que ele de si próprio diz.

Os novos e complexos desenvolvimentos da teoria da argumentação, propostos por Ducrot em colaboração com Marion Carel, cruzam, justamente, a argumentação com a noção de polifonia. Também na Análise do Discurso e na Pragmática se desenvolve investigação na senda das aperturas de Ducrot. Basta lembrar a riqueza teórica da ScaPoLine, a teoria escandinava da polifonia, devedora dos ensinamentos de Ducrot, ou o que a teoria da enunciação deve às figuras que configuram o seu conceito de polifonia.

Nascido em Paris, em 1930, Ducrot estudou Filosofia na École Normale Supérieure. Ensinou Filosofia antes de ter entrado no Centre National de la Recherche Scientifique, onde foi investigador e completou a sua formação com estudos de lógica matemática. Foi, posteriormente, professor na École des Hautes Études en Sciences Sociales. O seu pensamento e a sua obra beneficiaram, obviamente, da vastíssima formação que teve nas áreas referidas.

Não por acaso, seria Óscar Lopes, também ele com sólida formação, simultaneamente, em filosofia e em matemática, além de em estudos clássicos, a emprestar-me *Les échelles argumentatives* e a sugerir-me a leitura de *Les Mots du discours* e de *Le Dit et le dire*, nesses longínquos primeiros anos da década de oitenta, em que foi meu professor e, depois, orientador de Mestrado. Foram as suas lições que me apresentaram Ducrot e que me sugeriram a utilidade das suas abordagens para o estudo dos “operadores de agulhagem comunicativa” de que me ocupei: “ora”, “cá” e “lá”. O fascínio pela área acompanha-me desde então. Mas

todos nós, linguistas, somos devedores, em maior ou menor grau, dos ensinamentos estimulantes de Ducrot. Por isso lhe prestamos, aqui, homenagem.

Referências

- Anscombre, J.-C., & Ducrot, O. (1983). *L'argumentation dans la langue*. Mardaga.
- Anscombre, J.-C., & Ducrot, O. (1976). L'argumentation dans la langue. *Langages*, 42, 5-27. <https://doi.org/10.3406/lgge.1976.2306>
- Bakhtine, M. (1998). *La Poétique de Dostoievski*. Folio.
- Bally, C. (1944). *Linguistique générale et linguistique française*. A Francke.
- Ducrot, O. (1980). *Dire et ne pas dire. Principes de sémantique linguistique* (2^a ed.). Hermann.
- Ducrot, O. (1980). *Le Dire et le Dit*. Minuit.
- Ducrot, O. (1980). *Les Échelles argumentatives*. Minuit.
- Ducrot, O. et al. (1980). *Les Mots du discours*. Minuit.
- Grice, H. P. (1975). Logic and conversation. *Syntax and semantics*, 3, 41-58.
- Levinson, S. (1983). *Pragmatics*. Cambridge University Press.