

Processamento de relações anafóricas com sujeitos omitidos em Português Europeu*

Armanda Costa, Gabriela Matos & Paula Luegi

Faculdade de Letras da Universidade de Lisboa

Centro de Linguística da Universidade de Lisboa

The current study deals with referential chains' processing in finite clauses in European Portuguese, involving different types of subject gaps: *pro*, wh-variable and NP-copy. Since conditions of economy guide syntactic processing, we raised the hypothesis that the less referential is the content of the null element and the more minimal is the distance between the antecedent and the gap, the lower are the costs of the cognitive processing. We conducted an eye-tracking reading experiment and we measured the reading time per character. The results obtained show that the online processing of coreferential *pro* was costlier than the processing of wh-variable and NP-copy.

Keywords: Reading, referential chains' processing, null pronouns, WH- and NP-gaps, principles of economy.

Palavras-chave: Leitura, processamento de cadeias referenciais, pronomes nulos, WH e NP gaps, princípios de economia.

1. O processamento de relações anafóricas

No processamento da linguagem verbal, considerando compreensão e produção, há uma sucessão de operações linguísticas que se realizam sob limitações do sistema cognitivo que suporta os usos da linguagem e sob condições criadas pela gramática da língua. Uma situação particularmente interessante no processamento em tempo real, isto é, aquele que ocorre durante o acesso e integração do *input* linguístico de forma incremental, é a que se refere à resolução de relações anafóricas.

* O estudo conducente à elaboração deste trabalho foi desenvolvido no âmbito dos projectos PTDC/LIN/67854/2006 e PTDC/LIN/66202/2006, financiados pela Fundação para a Ciência e Tecnologia. Agradecemos à audiência do *XXV Encontro da Associação Portuguesa de Linguística* e a um revisor os comentários feitos a versões prévias do presente texto.

As relações anafóricas com pronomes não-reflexos ou recíprocos envolvem mais de um domínio frásico. Os pronomes são tipicamente categorias com potencial referencial intrínseco (1)¹, que podem, no entanto, participar em relações anafóricas, em que se estabelece uma relação de co-referência entre termos descontínuos (2). Esta relação pode ser afectada por múltiplos factores relevando quer dos princípios gramaticais que definem condições em que um pronome pode ser identificado por um antecedente, quer por condições do discurso que favorecem, ou não, a sua associação a um antecedente.

(1) *Nós* compramos o jornal.

(2) Eles viram as *crianças* quando *elas* entraram na biblioteca.

No caso do processamento de pronomes participando em cadeias de co-referência, impõe-se ao processador que, com rapidez e eficácia, encontre uma expressão referencial em condições estruturais e discursivas adequadas à identificação do conteúdo referencial do pronome. Tal operação deve fazer-se com custos mínimos considerando o sistema computacional da gramática e as limitações cognitivas do sistema humano para o processamento de frases.

Decorrente da investigação no domínio específico do processamento de relações anafóricas, têm sido identificados os vários tipos de categorias sintácticas envolvidas no processamento, assim como as condições que contribuem para uma variação de custos cognitivos na atribuição de um antecedente a um elemento com valor anafórico (Fodor, 1993; Featherston, 2001; Nicol e Swinney, 2003). As propriedades linguísticas do elemento anafórico em termos da informação intrínseca que disponibiliza ao processador são de crucial importância para a rapidez e a eficácia dessa associação. Devem considerar-se, neste caso, traços morfológicos de género, de número e até de caso que restringem possibilidades de associação do pronominal a uma expressão referencial numa dada posição gramatical. Contudo, no caso do Português Europeu, como língua românica que aceita sujeitos nulos, este tipo de informação linguística pode ser veiculado, não pelo pronominal nulo, mas pelas marcas de concordância da flexão verbal.

(3) a. Eles viram a *criança* quando *pro* entraram na biblioteca.

b. Ele viu as *crianças* quando *pro* entraram na biblioteca.

Relativamente ao potencial antecedente para um pronome, aspectos linguísticos e discursivos são identificados como podendo afectar a eficácia do processamento co-referencial. Não raramente, um pronome pode ir buscar o seu conteúdo referencial a mais do que uma entidade nominal que o antecede (4).

¹ Exclui-se, obviamente, destes casos os pronomes expletivos e os pronomes que, em línguas como o inglês, funcionam como variáveis ligadas.

(4) A *Ana* apresentou a *Maria* à *Joana* embora *ela* estivesse com pressa.

Haverá sobrecarga cognitiva com aumento de custos de processamento, quando mais do que um antecedente se perfila como candidato para um dado elemento anafórico; tal atrasa a decisão do processador tornando-a menos automática. Um factor decisivo nestas condições terá a ver com o grau de activação em memória de trabalho do possível antecedente, o que lhe vai conferir níveis de proeminência variáveis. A proeminência será uma consequência de propriedades estruturais (ocupar a posição de Sujeito frásico), (5a), discursivas (ser tópico discursivo ou frásico (5b), ser primeira referência no discurso (5a,b). Estas propriedades em convergência contribuem para o pico mais elevado de activação de um possível antecedente para um pronome.

(5) a. *Os rapazes* encontraram os amigos porque *pro* entraram na biblioteca.

b. *Aos amigos*, eles oferecem presentes, porque *pro*(*eles*) merecem.

Se se pensar na relação entre termos da cadeia referencial, serão propriedades relevantes aquelas que fazem variar a distância entre termos. A distância poderá ser considerada em termos de linearidade, entendendo-se por tal a distância medida em quantidade de itens lexicais com sobrecarga de informação que terá o seu peso no armazenamento em memória, ou em termos de complexidade estrutural do material interveniente. Comparando a resolução de *pro* na oração adverbial temporal, põe-se a hipótese de que em (6b) o *pro* acarrete mais custos de processamento do que em (6a), mais ainda devido à ocorrência da relativa parentética.

(6) a. *Os rapazes* encontraram os amigos na véspera do exame, quando *pro* entraram na biblioteca.

b. *Os rapazes* encontraram os amigos, de quem *pro* muito gostavam, quando *pro* entraram na biblioteca.

Para além de cadeias co-referenciais, colocam-se também como condições interessantes para o processamento aquelas em que temos um único elemento referencial distribuído por dois objectos sintácticos cuja interpretação obriga ao estabelecimento de relações dependenciais. Estão neste caso, construções em que, por operações dentro de um mesmo domínio sintáctico, há movimento de um constituinte para uma posição não básica. Neste caso, o processador tem de identificar o lugar na estrutura de onde foi movido o constituinte e de identificar o lugar vazio na estrutura que terá de ser ligado a uma expressão referencial que ocupa uma posição estrutural derivada, não básica.

(7) a. O rapaz *a quem* eles telefonaram [-] foi o Pedro.

b. *Aos amigos*, eles telefonam [-] frequentemente.

Quando se tenta identificar as fontes de informação que apoiam o processador e os princípios de que se serve para resolver cadeias anafóricas, gramática e processamento são convocados em simultâneo por propriedades que os unificam. Decorrendo o processamento em tempo real (e estamos a referir-nos apenas à modalidade de compreensão) em condições de recursos limitados de memória, de automatização de operações, de tomada de decisões com informação escassa tendo em vista uma estruturação e interpretação bem sucedidas, princípios que garantam uma economia de procedimentos são essenciais. E estes são contemplados quer nos modelos de processamento de frases quer no modelo de gramática de natureza computacional.

Em termos gerais, poderá dizer-se que o processador será guiado pela preferência (i) por formas anafóricas mínimas para retomar antecedentes proeminentes (por exemplo, pronomes nulos para retomar sujeitos frásicos); (ii) por formas máximas para retomar antecedentes estrutural ou discursivamente não proeminentes (por exemplo, expressões nominais descritivas para retomar antecedentes recuados no discurso ou em condições em que não são tópicos); (iii) por cadeias estruturais económicas (por exemplo, ligações sintácticas locais envolvendo menos passos computacionais).

Consideremos dois tipos de estruturas em que no *input* linguístico há pistas de que o processador se serve para resolver tão eficazmente quanto possível a estruturação e a interpretação da frase.

Tomemos o exemplo de frases interrogativas ou relativas em que há constituintes movidos para a periferia esquerda da frase (constituintes WH-), que funcionam como sinais de alerta (*fillers*) para o *parser* (analizador sintáctico), antecipando a existência de subseqüentes *gaps* (relação *filler/gap*):

- (8) a. Pergunto-me *que* livros o João terá lido [-] antes de fazer o exame.
 b. *Que* leu o João [-] antes de ir para exame?
 c. Estes são os livros que o João leu [-] antes de ir para o exame.
 d. Estes são os livros que a Maria disse ao Pedro que o João leu [-] antes de ir para o exame.

Nestas condições, Frazier (1987), Frazier & Clifton (1996), propõem a actuação de um princípio que designam por *Filler-gap strategy* e que determina que:

“Once an element of a category XP is identified as moved from its argument position, then posit a corresponding empty XP category as soon as the grammar of the language allows you to do so.”

Assim, sempre que um *filler* é encontrado, automaticamente é predita a existência de um lugar vazio na estrutura (um *gap*). O *filler* fica activo em memória de trabalho e o sistema de processamento procura activamente o *gap* onde esse *filler* deve ser ligado.

Consideremos agora construções em que nada anuncia a existência de movimento nem a necessidade de ligar objectos sintácticos descontínuos e dependentes entre si (vejam-se os exemplos em (9) correspondentes aos propostos para o italiano por De Vincenzi, 1991:16) onde estão assinaladas as leituras preferenciais:

- (9) a. No mercado chamou [o vendedor]_{SU/OD}...
 b. No mercado *pro* chamou [o vendedor]_{OD} para lhe pedir um desconto.
 c. No mercado chamou [o vendedor]_{SU} [-]_{OD} para lhe oferecer um desconto.

Em (9a), a frase pode ser interpretada como sendo constituída por um verbo transitivo cujo Objecto está explícito e cujo Sujeito é um pronominal nulo, ou como uma construção intransitiva com inversão de Sujeito. Desambiguando a interpretação, em (9b) há uma construção transitiva *SproVO*, e em (9c) uma construção com inversão do sujeito (VS).

Para este tipo de construções com verbos potencialmente transitivos ou intransitivos, com construções interrogativas com ambiguidade entre extracção de Sujeito ou de Objecto, com construções inacusativas, De Vincenzi propõe o seu *Minimal Chain Principle*, que propõe o seguinte:

“*Avoid postulating unnecessary chain members at surface structure, but do not delay required chain members*” (De Vincenzi, 1991:13).

O sistema de processamento, seguindo critérios de economia, deverá preferir cadeias mais simples, mais locais, com menos movimento, por isso mais económicas. Isto é, o *parser* manifesta sensibilidade à complexidade da estrutura.

Consideradas unidades maiores do que a frase de natureza discursiva, emergem teorias que estabelecem relações entre a proeminência do Antecedente e a forma da expressão anafórica. *The Accessibility Marking Scale* (Ariel, 1996) propõe uma relação de correlação inversa entre o elemento anafórico em termos da sua informatividade, do seu peso fonológico e da sua estrutura e o nível de activação do Antecedente: *gaps* (*pro* e variáveis) ou pronomes (clíticos e não-clíticos) funcionam como marcadores de alta acessibilidade do antecedente (Sujeito ou Tópico recentes), expressões nominais plenas, mais ou menos elaboradas (incluindo um complemento ou um modificador) são marcadores de baixa acessibilidade do antecedente (não tópicos, constituintes não Sujeito).

A Teoria da Centralização (*Centering Theory*), proposta por Grosz, Joshi & Weinstein (1995), explica relações de coerência local no discurso. A entidade mais proeminente do discurso é aquela que desempenha a função de Sujeito e ocupa a posição estrutural mais alta; como tal deve ser retomada por um pronome, nunca por uma expressão nominal repetida, sob risco de se criar uma condição desfavorável ao estabelecimento de co-referência. Veja-se que relativamente a (10), (10a) constitui sequência adequada em que o Sujeito e Tópico da oração precedente são retomados por um pronome. Já em (10b), a repetição do nome gera diminuição da aceitabilidade. Em (10c), por ter havido transição de tópico, a repetição do SN pleno não gera inaceitabilidade.

- (10) *A Paula* deu um periquito à Carla.
 a. *Ela* disse à Carla que os periquitos são vaidosos.

- b. ??A Paula disse-lhe que os periquitos são vaidosos.
- c. A Carla agradeceu à Paula a oferta.

Sintetizando, poderá afirmar-se que no processamento de cadeias com elementos anafóricos:

(i) Teorias de hierarquia e de acessibilidade prevêm que as estratégias de processamento são motivadas pela forma fonética, morfológica e sintáctica da expressão anafórica que funciona como pista para a busca de um antecedente.

(ii) A cabeça da cadeia referencial pode ter diferentes níveis de activação de acordo com a sua posição estrutural, o seu papel temático, a sua posição na frase ou o seu estatuto pragmático como tópico.

(iii) Quando uma expressão nominal é Tópico, Sujeito e é Primeira Referência possui um conjunto de condições que a promovem a melhor Antecedente de uma expressão anafórica que ocorra numa posição mais baixa.

2. Aspectos sintácticos no processamento de relações anafóricas

No Programa Minimalista actual, a gramática é vista como um sistema computacional que, operando sobre objectos sintácticos, deriva representações interpretadas pelas componentes fonológica e semântica, as quais constituem níveis de interface para os sistemas periféricos acústico-articulatório e o conceptual-intencional, externos à gramática.

Esse sistema computacional funciona de forma altamente restritiva e económica, sendo apenas legítimos as categorias, as operações e os princípios que recebem a sua motivação dos sistemas periféricos externos.

Princípios de economia guiam o seu funcionamento, visando otimizar a derivação das expressões linguísticas. Assim, as operações gramaticais são guiadas por um princípio de localidade que restringe a sua actuação aos elementos relevantes estruturalmente mais próximos, em domínios locais, produzindo as derivações mais curtas (Chomsky, 2001, 2004, 2008) e as derivações processam-se em conformidade com o *Earliness Principle* (Pesetsky, 1989), que determina que as computações sejam realizadas tão cedo quanto possível (Chomsky, 2001:15). Deste modo, logo que um domínio predicativo (ν P) ou proposicional CP) completo, i.e. uma *fase*, não é requerido para a computação da fase seguinte, ele é *transferido* para as interfaces fonológica e semântica para ser, respectivamente, soletrado [*spelled out*] e interpretado (Chomsky, 2004, 2008).

Neste sistema, as relações anafóricas são em grande parte captadas pelos princípios da Teoria da Ligação, que guiam a interpretação das expressões nominais no nível de representação semântica.

No presente trabalho, centrar-nos-emos nas propriedades das categorias vazias nominais que ocorrem em posição de sujeito das frases finitas: *pro*, variável-WH e cópia de movimento-NP. As relações anafóricas em que estes constituintes omitidos participam decorrem da natureza do constituinte nominal e do tipo de derivação envolvido.

Pronominal nulo (pro)

O pronome nulo é gerado na base e inserido na posição argumental com a qual está relacionado, obedecendo ao Princípio B da Teoria da Ligação (i.e., não pode ser ligado num domínio de ligação local). Ocorre em posição de Sujeito de orações raiz (11) e subordinadas finitas em línguas de Sujeito Nulo (13), (14) e, ainda, em orações de infinitivo flexionado (12). Do ponto de vista referencial, apresenta conteúdo independente definido (11) ou arbitrário (12), podendo também ocorrer como um expletivo, como na primeira ocorrência em (13). Numa relação de co-referência, depende referencialmente de um antecedente e forma com ele uma cadeia referencial complexa (14):

(11) *pro* Lemos esse livro.

(12) É escusado *pro arb* dizê-lo.

(13) *pro* Parece que *pro* lemos esse livro.

(14) O homem levantou-se do degrau quando *pro* ouviu correr os ferrolhos.

Variáveis Wh

As variáveis WH são categorias nulas resultantes do movimento de um sintagma-WH para a periferia esquerda da frase (Movimento-A'). Ocorrem numa cadeia que têm como elos o sintagma-WH, em posição não argumental, e o seu *gap*/cópia em posição argumental, onde o caso é verificado; o seu conteúdo depende da ligação ao constituinte A'-movido e da posição argumental que ocupa. Em termos da Teoria da Ligação são caracterizados como Expressões-R(eferenciais), i.e., expressões que não podem ter um constituinte em posição argumental que as ligue (Princípio C):

(15) O homem $[[WH- que]_{TP} [WH-] queria um barco.]]$

Cópias de NP

As cópias de NP são categorias nulas resultantes do movimento de uma categoria nominal para uma posição argumental (Movimento-A), onde o seu caso é verificado. Ocorrem em cadeias encabeçadas por NP movidos; o seu conteúdo depende da ligação ao constituinte A-movido. Obedecem ao princípio A da Teoria da Ligação, ou seja são anáforas (têm de ser localmente A-ligadas):

(16) *O homem* foi recebido $[NP-]$ pelo rei.

3. Trabalho experimental

Na seqüência de trabalho sobre o processamento de relações anafóricas em Português Europeu (Costa, Faria e Matos, 1998; Costa, Faria e Kail, 2004; Costa, 2005; Costa e Luegi, 2009), são objectivos do estudo que agora se apresenta avaliar custos de processamento de cadeias anafóricas com propriedades gramaticais distintas, em função

do tipo de expressão anafórica (pronomes nulos, variáveis WH ou cópias de NP), e da complexidade das estruturas sintáticas que albergam as expressões anafóricas.

Partindo do pressuposto de que condições de minimalidade guiam o processamento sintático das relações anafóricas, foram formuladas as seguintes hipóteses: (i) quanto menor for o valor referencial do elemento anafórico nulo e mais curta, em termos de estrutura hierárquica, for a distância entre o antecedente e o constituinte não realizado, mais baixos são os custos de processamento cognitivo; (ii) considerando *pro*, WH- e NP-gaps, prevê-se que os custos de processamento sejam mais elevados nos contextos onde *pro* ocorre.

3.1. Desenho experimental

Participantes

Participaram neste estudo 20 alunos de graduação e pós-graduação da Faculdade de Letras da Universidade de Lisboa. Todos os participantes eram falantes nativos de Português Europeu com hábitos regulares de leitura.

Tarefa experimental

A tarefa consistiu na leitura de um excerto do texto *O Conto da Ilha Desconhecida*, de José Saramago. O texto foi apresentado no ecrã de um computador dividido por 10 slides, sendo a leitura auto-controlada, slide a slide. Antes de dar início à tarefa, os participantes foram informados sobre o protocolo a seguir, que incluía o preenchimento de um questionário de compreensão após a leitura de todos os slides.

Procedimento

Durante a leitura do texto, os comportamentos oculares dos participantes foram registados com um sistema de *eye-tracking*, modelo R6-HS da ASL, a uma velocidade de 240 Hz. Em cada registo individual, após a calibração do equipamento, foi apresentado um pequeno texto para treino de leitura em situação experimental.

Dados experimentais

No excerto seleccionado identificaram-se as diferentes estruturas que continham *pro*, Wh-gap ou NP-gap, como exemplificado de (a) a (e).

(a) *este chamava o terceiro, que mandava o primeiro-ajudante, que por sua vez mandava o segundo*

este chamava o terceiro_P, [_{CP} que_i [_{TP} [WH-]_i mandava o primeiro-ajudante, [_{CP} que_j [_{Parentética} por sua vez] [_{TP} [WH-]_j mandava o segundo]]]

(b) *até chegar à mulher da limpeza, a qual, não tendo ninguém em quem mandar, entreabria a porta e perguntava*

até *pro* chegar à mulher da limpeza_i, [a qual_i [_{Parentética} não tendo *pro*_i ninguém [_{CP} em quem [*pro*_i mandar]]] [_{TP} [WH-]_i entreabria a porta e [[WH]_i perguntava]].

(c) *ao primeiro-secretário, o qual, escusado seria dizer, passava a encomenda ao segundo-secretário*

ao primeiro-secretário_i, [_{Parentética} o qual_i, [_{Parentética} *pro* escusado seria dizer], [_{TP} [WH-]_i passava a encomenda ao segundo-secretário_j]]

(d) *até chegar à mulher da limpeza, que despachava sim ou não conforme estivesse de maré*

até *pro* chegar à mulher da limpeza_i, [_{CP} que_i [_{TP} [WH-]_i despachava sim ou não conforme *pro*_i estivesse de maré]

(e) *O homem que queria um barco levantou-se do degrau da porta quando começou a ouvir correr os ferrolhos, enrolou a manta e pôs à espera.*

[_{TP}[O homem [_{CP} que_i [_{TP} [WH-]_i queria um barco]]]_i [_{TP1} [NP-]_j levantou-se do degrau da porta [quando *pro* começou a ouvir correr os ferrolhos]], [_{TP2} [NP]_j enrolou a manta] e [_{TP3} [NP-]_j pôs à espera]].

Os dados experimentais apresentam diferentes níveis de complexidade estrutural.

As cadeias-WH, apesar de serem todas locais, restringindo a uma só Fase de CP, apresentam propriedades distintas. Algumas apresentam contiguidade entre a expressão WH e o seu lugar de extracção, como acontece em (a) na oração relativa *que mandava o primeiro-ajudante*. Porém, outras exibem interrupções produzidas por expressões parentéticas; estas exibem entre si diferentes graus de complexidade estrutural: em (a), há um sintagma interposto (*por sua vez*), enquanto que em (b) há uma oração gerundiva comportando uma oração relativa (*não tendo ninguém em quem mandar*). Adicionalmente, uma das cadeias-WH manifesta, ainda, efeitos de extracção simultânea (*across-the board*) do constituinte relativizado, *a qual_i (...) [WH-]_i entreabria a porta [WH-]_i e perguntava*.

Quanto às cadeias NP, limitam-se a uma Fase de CP e apresentam igualmente efeitos de extracção simultânea, *O homem... [NP-]levantou-se do degrau da porta quando começou a ouvir correr os ferrolhos, [NP-]enrolou a manta e [NP-] pôs à espera*.

As cadeias de co-referência com *pro*, por seu turno, envolvem pelo menos duas Fases de CP.

3.2. Análise de dados

Em cada um dos cinco fragmentos foram criadas várias regiões de análise, delimitadas pela barra vertical (|), com exemplificado em (17). A fronteira de cada região coincide com fronteiras oracionais que, por sua vez, se iniciam por uma expressão anafórica nula na posição de sujeito frásico e tópico não marcado.

(17) O homem que queria um barco | levantou-se do degrau da porta | quando começou a ouvir correr os ferrolhos, | enrolou a manta | e pôs à espera.

Começámos por analisar, como variáveis dependentes, os seguintes valores: (i) *Tempo Total de Leitura*, soma da duração de todas as fixações realizadas numa região; (ii) *Número de Fixações*, número total de fixações realizadas numa região; (iii) *Duração Média das Fixações*, duração média das fixações realizadas numa região; (iv) *Tempo de Leitura por Caracter*, tempo total de leitura da região dividido pelo seu número de caracteres. Nos resultados que agora apresentamos, apenas nos ativemos ao *Tempo de Leitura por Caracter*, medida que usámos para poder comparar regiões que não possuem a mesma extensão (consequência do uso de texto não manipulado).

Apresentamos na Tabela 1 os valores médios do tempo de leitura por caracter nas diferentes regiões de análise (o número das regiões que contêm cada um dos diferentes tipos de expressão anafórica está também representado na tabela, na coluna N.º de Regiões).

WH-gap	4	49 ms	51 ms	49 ms
WH-gap (parentética)	3	53 ms	45 ms	
NP-gap	3	45 ms	56 ms	
pro	4	61 ms		
pro arbitrário	1	50 ms		

Tabela 1: Tempos médios de leitura por caracter (T/Car), em milésimos de segundo (ms), das diferentes regiões que contêm expressões anafóricas.

Pela análise do Gráfico 1, podemos verificar que existem diferenças no tempo de leitura por caracter de algumas regiões que contêm elementos anafóricos distintos.

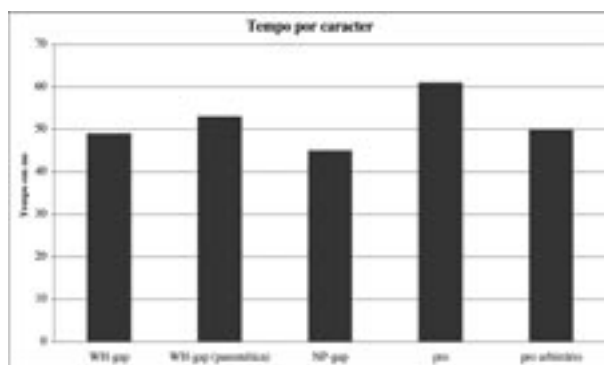


Gráfico 1: Tempo de leitura por caracter (T/Car), em milésimos de segundo (ms), das diferentes regiões com elementos anafóricos.

Numa primeira análise, contrastando os valores médios dos três factores linguísticos: *pro* (excluindo *pro* arbitrário por haver apenas uma ocorrência), WH- (considerando a média de todas as ocorrências) e de NP-, verificam-se diferenças significativas entre o valor das médias obtidas, a um nível de significância de 0,05 ($F(2,57) = 7,779$; $p = 0,001$). O teste *Post Hoc* de *Games-Howell* revela que há diferenças estatisticamente significativas entre as condições WH- ($M=0,047$) e *pro* ($M=0,060$) ($p=0,021$) e entre NP ($M=0,045$) e *pro* ($M=0,060$) ($p=0,003$).

Contrastando as condições em que há *gaps* gerados por movimento e as condições em que há *pro*, pode verificar-se que a média do T/Car nas regiões com WH-*gap* e NP-*gap* é de 49 ms, claramente mais baixo do que o tempo médio em contextos com *pro* co-referente, 61 ms (note-se que estamos a excluir o valor médio de *pro* arbitrário, de que só temos uma ocorrência com o valor médio de 50ms).

No contraste entre as regiões com variáveis WH- e com cópias de NP-, os valores mais baixos encontram-se nas cópias de NP, com o valor médio de 45 ms contra 51 ms nas regiões WH-*gap*.

Atendendo às condições de maior ou menor complexidade sintáctica em condições de cadeias WH-, considerámos condições em que há contiguidade entre o constituinte movido, assinalado pelo sintagma WH- e o seu lugar de extracção, que é sempre o de Sujeito frásico. Comparados os valores obtidos nos dois tipos de condições, verifica-se que ao aumento de complexidade estrutural se associa um aumento dos tempos de leitura: quando a expressão WH- é contígua ao lugar de extracção do constituinte temos o valor de 49 ms, quando há descontinuidade o valor é de 53 ms. Trata-se de uma diferença não negligenciável para futuros trabalhos no domínio.

4. Conclusões

Propusemo-nos neste trabalho testar empiricamente custos de processamento de elementos anafóricos sem realização fonética em posição de sujeito de orações subordinadas e coordenadas finitas. Estes elementos participam em cadeias anafóricas e apresentam propriedades sintácticas distintas. Para a sua resolução no processamento em tempo real, formulámos hipóteses que aqui retomamos:

- quanto menor for o valor referencial do elemento anafórico nulo e mais curta, em termos de estrutura hierárquica, for a distância entre o antecedente e o constituinte não realizado, mais baixos são os custos de processamento cognitivo;
- considerando *pro*, WH- e NP-*gaps*, prevê-se que os custos de processamento sejam mais elevados nos contextos onde *pro* ocorre.

Os dados obtidos permitem-nos confirmar as nossas hipóteses. Relações anafóricas com *pro* co-referencial são mais difíceis de processar do que *gaps* resultantes de movimento. Dos três tipos de categorias sintácticas testadas, *pro*, que nas condições

analisadas participa numa cadeia de co-referência, é o único pronominal e, como tal, tem conteúdo referencial que pode ser recuperável numa primeira análise pela flexão verbal e, numa segunda etapa, pela recuperação do seu antecedente, envolvendo assim duas fases computacionais de CP. Além disso, o processamento do pronome implica a consulta de indicadores discursivos. Poderá dizer-se que o processamento de *pro* impõe custos acrescidos visto que gramática e discurso têm de ser consultados para identificação do seu conteúdo referencial.

Contextos que envolvem *gaps* por movimento apresentam custos mais reduzidos do que aqueles em que *pro* ocorre. À partida, contamos com uma propriedade sintáctica que distingue crucialmente estes objectos sintácticos de *pro*: o seu conteúdo referencial depende exclusivamente de condições gramaticais para poderem ser ligadas e identificadas, no seu domínio local. Esta propriedade pode explicar os valores médios significativamente mais baixos na leitura das regiões onde ocorrem, comparativamente às que integram *pro*. Poderemos então concluir que sujeitos anafóricos que durante o processamento sintáctico são resolvidos exclusivamente no âmbito das restrições gramaticais constituem condições favoráveis à actuação do analisador sintáctico (*parser*). A não opcionalidade na busca de antecedente, a inexistência de outras fontes de informação linguística que não a sintáctica habilitam o *parser* a proceder com mais rapidez e eficácia.

Quando comparados os resultados obtidos na leitura de regiões onde ocorrem os dois tipos de *gaps* de movimento, verificam-se diferenças que dão vantagem às cópias de NP. WH- e NP-*gaps* partilham propriedades, dado ambos serem resultantes de movimento e ligados pelo constituinte movido que está localizado na mesma fase de CP. Consideradas as propriedades sintácticas das condições em que ocorrem e os processos implicados no seu processamento, dir-se-ia que os *gaps* de WH- deveriam apresentar condições mais favoráveis à sua resolução do que os *gaps* de NP. Na verdade, o *filler*-WH funciona como um alerta para o *parser* procurar o *gap* (*Active Filler Strategy*), condição inexistente em caso de *gap* de NP em que apenas a grelha argumental não abertamente preenchida do verbo assinala o movimento do constituinte. Contudo, NP-*gaps* elicitam um comportamento mais automático, mais rápido. Interpretamos estes resultados como um indicador de um possível efeito de menores custos computacionais no processamento de cópia de NP. Este, comparativamente ao *gap*-WH, envolve menos custos computacionais, considerada a derivação da estrutura e tendo em vista, em conformidade com a caracterização proposta pela Teoria da Ligação, o estatuto referencial de cada um destes elementos: enquanto anáfora, a cópia de NP é desprovida de conteúdo referencial intrínseco; pelo contrário, a variável WH, é uma expressão argumental com conteúdo referencial pleno, um expressão-R.

Em suma, relações anafóricas envolvendo Sujeitos omitidos não são uniformes no que diz respeito aos custos de processamento. O seu estatuto sintáctico, o facto de a sua resolução ser feita exclusivamente no âmbito de restrições gramaticais ou depender de informação discursiva, o facto de participarem em estruturas com derivações mais ou menos complexas, tudo vai ter impacto nos custos do seu processamento.

Retomados os princípios de *parsing* acima referidos e a escala de acessibilidade do antecedente proposta por Ariel, poderá concluir-se que o *parser* se guia por princípios de economia em termos cognitivos e computacionais e que qualquer escala de hierarquia para expressões referenciais que tente explicar o processamento anafórico tem de incluir outros elementos nulos para além de *pro*, devendo considerar que estes elementos não têm todos os mesmos custos de processamento.

Referências bibliográficas

- Ariel, Mira (1996) Referring expressions and +/- coreference distinction. In T. Fretheim & J. Gundel (eds.). *Reference and referent accessibility*. Amsterdam: John Benjamins Publishing Company, pp. 13-35.
- Chomsky, Noam (2001) Derivation by phase. In Michael. Kenstowicz, *Ken Hale. A life in language*. Cambridge, Massachusetts: The MIT Press, pp. 1-52.
- Chomsky, Noam (2004) Beyond explanatory adequacy. In Adriana Beletti, (ed.). *Structures and beyond: The cartography of syntactic structures*, vol. 3. Oxford: Oxford University Press, pp. 104-131.
- Chomsky, Noam (2008) On Phases. In Robert Freidin, Carlos Otero & Maria Luisa Zubizarreta, *Foundational Issues in Linguistic Theory*. Cambridge, Massachusetts: The MIT Press, pp. 133-166.
- Costa, Armanda (2005) *Processamento de frases em Português Europeu*. Lisboa: Fundação Calouste Gulbenkian.
- Costa, Armanda & Paula Luegi (2009) Complexidade linguística e processamento referencial. In Dermeval da Hora (ed.), *Anais do VI Congresso Internacional da ABRALIN* (CD-ROM). João Pessoa.
- Costa, Armanda, Isabel Hub Faria & Gabriela Matos (1998) Competitive information sources in referential ambiguity resolution. *Proceedings of 5th International Congress of the ISAPL*. Porto, pp. 173-187.
- Costa, Armanda, Isabel Hub Faria & Michèle Kail (2004) Semantic and Syntactic Cues' Interaction on Pronoun Resolution in European Portuguese. In António Branco, Tony McEnery & Thomas Mitkov (eds.) DAARC 2004, *Proceedings of 5th Discourse Anaphora Resolution Colloquium*. Lisboa: Ed. Colibri, pp. 45-50.
- De Vincenzi, Marica (1991) *Syntactic Parsing Strategies in Italian*. London: Kluwer Academic Publishers.
- Featherston, Sam (2001) *Empty categories in sentence processing*. Amsterdam: John Benjamins Publishing Company.
- Fodor, Janet (1993) Processing empty categories: a question of visibility. In Gerry Altmann & Richard Shillcock (eds.). *Cognitive models of speech processing: the Second Sperlonga Meeting*. Hillsdale: Lawrence Erlbaum Associates, Publishers, pp. 351-400.
- Frazier, Lyn & Charles Clifton (1996) *Construal*. Cambridge: A Bradford Book.

- Frazier, Lyn (1987) Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5, pp. 519-559.
- Grosz, Barbara, Aravind Joshi & Scott Weinstein (1995) Centering: A Framework for Modeling the Local Coherence of Discourse. *Computation Linguistics*, 21 (2), pp. 203-225.
- Mateus, Maria Helena, Ana Maria Brito, Inês Duarte, Isabel Hub Faria, Sónia Frota, Gabriela Matos, Fátima Oliveira, Marina Vigário, Alina Villalva (2003) *Gramática da Língua Portuguesa*. Lisboa: Caminho.
- Nicol, Janet & David Swinney (2003). The psycholinguistics of anaphora. In Andrew Barss (Ed.). *Anaphora: a reference guide*. Oxford: Blackwell Publishers, pp. 72-104.