

ALGUNS FATORES RELEVANTES NO PROCESSAMENTO DA CORREFERÊNCIA INTERSENTENCIAL

Juciane Nóbrega Lima é Mestre em Linguística pela UFPB.

E-mail: jucy.nobrega@gmail.com

Flávia Gonçalves Calaca de Souza é Doutoranda em Linguística na UFPB.

E-mail: flavia.calaca@hotmail.com

Resumo

O presente artigo tem como objetivo apresentar um panorama geral de alguns fatores que podem atuar no processamento da correferência. Tais como a Teoria da Centralização; Proeminência Sintática; Hipótese da Carga Informacional; Paralelismo Estrutural; e Foco. O processamento da correferência procura investigar como se dão as relações entre os modelos abstratos mentais e o sistema de acesso a esta representação em tempo real.

Abstract

This paper intends to present an overview of some factors that may influence the processing of coreference. Including the Centering Theory; Syntactic prominence; Informational Load Hypothesis; Structural parallelism; and Focus. The coreference processing investigates how the relationship between mental abstract models and the system of access to this representation happens in real time.

1) Introdução

Para que um discurso tenha uma estrutura coerente, deve-se haver uma integração entre seus enunciados. Desta forma, o modo como uma entidade se refere a um elemento já mencionado no discurso influencia na integração desses enunciados. Este tipo de integração é denominado correferência, que pode ser estabelecida de várias formas anafóricas: através de pronomes, nomes repetidos, categoria vazia, hiperônimos, hipônimos, etc. A correferência acontece mentalmente de forma dinâmica, sendo incluídos entidades e eventos estabelecendo relações que vão se atualizando a medida que as sentenças ou o próprio discurso transcorrem e as informações novas surgem e se desdobram.

Essas formas linguísticas correferenciais supracitadas podem aparecer em posição de sujeito ou de objeto, tanto no escopo da sentença (intrassentencial), quanto no escopo do discurso (intersentencial). As relações anafóricas intrassentenciais geralmente procuram observar como a Teoria da Ligação (*Binding Theory*) (CHOMSKY, 1981) influencia ou mesmo guia tais relações no processamento *on-line*. Já os estudos que observam as relações anafóricas intersentenciais procuram observar como fatores tanto estruturais e gramaticais (por exemplo: paralelismo estrutural, concordância), quanto fatores semântico-pragmáticos (por exemplo: tipo de retomada, saliência discursiva ou proeminência sintática, animacidade) podem influenciar no estabelecimento da correferência.

Neste artigo abordaremos os principais pressupostos e teorias que subjazem o processamento da correferência intersentencial. Primeiramente faremos uma explanação sobre a Teoria da Centralização, em seguida falaremos sobre a noção de Proeminência Sintática e da Hipótese da Carga Informacional expondo os principais estudos de processamento linguístico que investigam o fenômeno da Penalidade do Nome Repetido. Depois há uma parte que trata sobre o foco, apresentando os principais estudos na área do processamento linguístico, mostrando como tal estrutura influencia

no estabelecimento da correferência. E, por fim, apresentaremos alguns estudos que investigam o Paralelismo Estrutural, mostrando como esse efeito pode influenciar no processamento da correferência.

2) Teoria da Centralização (*Centering Theory*)

A Teoria da Centralização (*Centering Theory*) propõe que a estrutura do enunciado e a coerência discursiva local têm por objetivo “dirigir o foco de atenção para a escolha de uma expressão referencial e para a coerência dos enunciados dentro do segmento de um discurso” (Gross, Joshi & Weinstein, 1995, p. 204). Para esta teoria, a coerência está relacionada a elementos denominados centros, que podem ser centros anafóricos (*Cb*) (*backward-looking centers*), que possibilitam a ligação com um enunciado precedente, ou centros catafóricos (*Cf*) (*forward-looking centers*), que servem como potenciais centros anafóricos (*Cb*) do enunciado subsequente.¹ Assim, concluiu-se que o centro anafórico (*Cb*) deve ser realizado como um pronome, pois, dessa forma, o discurso se torna coerente.

Os membros do conjunto de centros catafóricos (*Cf*) são ordenados hierarquicamente de acordo com a proeminência de cada um dentro do enunciado. Essa proeminência determina a facilidade com que uma entidade em um enunciado pode servir como centro anafórico (*Cb*) do enunciado subsequente, ou seja, quanto mais proeminente o centro anafórico (*Cb*) estiver no discurso, mais facilmente será recuperado.

Desta forma, a teoria postula que a utilização de um pronome anafórico, sendo um centro anafórico (*Cb*) para designar uma entidade saliente, ou seja, em foco no discurso, traz um índice explícito, deixando o centro catafórico (*Cf*) mais ativo na memória de trabalho. No entanto, o emprego de uma expressão como um nome próprio, ou um SN definido, torna essa relação mais difícil de ser estabelecida, gerando assim um custo maior de tempo no processamento, uma vez que ao utilizar um nome o leitor/ouvinte pode interpretá-lo como um novo centro catafórico (*Cf*).

Essa predição da Teoria da Centralização foi verificada por Gordon et al. (1993), utilizando-se a técnica experimental da leitura automonitorada (*Self-paced reading*), e os resultados encontrados mostraram que, realmente, o uso de nomes repetidos torna a leitura mais lenta em comparação com o uso de pronomes, e esse tipo de efeito foi nomeado de Penalidade do Nome Repetido (*Repeated-Name Penalty*).

De acordo com Gordon et al. (1993), esta Penalidade ocorre quando há ausência de um pronome correferente ao antecedente que está em uma posição saliente no discurso, uma vez que os pronomes são considerados por esta teoria os veículos naturais para o estabelecimento da correferência, sendo utilizados como uma pista relevante de coerência local. Assim, a sua ausência induziria o leitor/ouvinte a uma falha de expectativa pela falta dessa pista, o que faria com que a compreensão fosse mais custosa.

3) Proeminência Sintática

¹ As traduções utilizadas foram as mesmas adotadas por Leitão (2005).

Nota-se que nos estudos já citados ainda não havia uma definição formal sobre o que seria a proeminência sintática e qual a sua real relevância para o processamento da correferência. Havia apenas uma referência à saliência do antecedente, sendo a posição de sujeito a mais saliente discursivamente.

Gordon & Hendrick (1998) propõem que a entidade considerada proeminente em um enunciado vai ser a que se encontra em uma posição de domínio na sentença dentro de uma relação de c-comando. Esta formalização baseia-se na noção de c-comando proposta por Reinhart (1981), em que a relação entre elementos correferenciais é estabelecida a partir de uma ordem hierárquica².

Com base nessa noção formal de proeminência sintática, têm-se encontrado evidências empíricas de que essa proeminência influencia o estabelecimento da correferência. Há diversos estudos que mostram que há Penalidade do Nome Repetido quando o antecedente é sujeito da sentença precedente, ou seja, na posição mais proeminente discursivamente, e não quando o antecedente está em posição de objeto. Por exemplo, os estudos em inglês Gordon et al. (1993); Kenissson & Gordon (1997); os estudos do mandarim Yang et al (2001) e os em francês Ermst (2007).

Em resumo, esse apanhado teórico sugere que diferentemente dos nomes próprios repetidos ou de sintagmas nominais plenos, os pronomes anafóricos têm uma tendência a designar entidades que são proeminentes nas representações do discurso local. Eles seriam, portanto, indicadores mais sensíveis de focalização de entidades do que os nomes repetidos ou sintagmas nominais plenos.

Um questionamento relevante, levantado em Gordon & Chan (1995), foi o da importância de um pronome para a coerência discursiva. Os autores questionam se essa coerência seria determinada pelo seu papel sintático, ou pelo seu papel temático dentro de uma sentença. A dissociação entre papel temático e papel sintático foi feita pelo uso de sentenças ativas e passivas em quatro experimentos. Em todos eles, o pronome foi lido mais rapidamente do que o nome repetido confirmando assim a Penalidade do Nome Repetido. Na maioria dos experimentos, esta penalidade ocorreu quando o antecedente estava com função sintática de sujeito gramatical. Entretanto, em um dos experimentos em que não havia na posição de sujeito gramatical uma entidade candidata possível a ser um centro anafórico (*Cb*), foi encontrada Penalidade do Nome Repetido para entidades na posição de objeto direto. Isso nos mostra que, quando não há na posição de sujeito gramatical um candidato disponível para se estabelecer a correferência e a exercer o papel de centro anafórico (*Cb*), outra posição pode ser utilizada para isso. Assim, os autores concluíram que a proeminência está relacionada a critérios especificamente sintáticos e não relações de papel temático.

Vimos até aqui que a Teoria da Centralização explica o conceito de proeminência sintática e o conceito de Penalidade do Nome Repetido, mas deixa a desejar quanto à abrangência da teoria relacionada a outras formas anafóricas, como os hiperônimos e hipônimos, por exemplo. Uma hipótese que se posiciona de maneira crítica em relação a essa limitação da Teoria da Centralização é a Hipótese da Carga

² Segundo a Teoria da Vinculação *A* c-comanda *B* se e somente se *B* é irmão de *A* ou filho do irmão de *A*.

Informacional (*Informational Load Hypothesis*) proposta por Almor (1999, 2000), que será discutida na próxima seção.

4) Hipótese da Carga Informacional (*Informational Load Hypothesis*)

A Hipótese da Carga Informacional (*Informational Load Hypothesis*), proposta por Almor (1999, 2000), oferece uma explicação diferente sobre o fenômeno da Penalidade do Nome Repetido, apesar desta hipótese não ser especificamente para explicar este fenômeno. O autor propõe que o processamento correferencial se dá através de uma relação entre a distância semântica entre os elementos correferenciais e o custo operacional da memória de trabalho, também propõe que um elemento anafórico precisa ter sua função discursiva justificada dentro de um contexto específico.

Em termos de custo de memória de trabalho, esta teoria afirma que pronomes são processados mais rapidamente do que nomes repetidos no estabelecimento da correferência em função da carga informacional que essas formas linguísticas contêm, ou seja, quanto maior a distância semântica entre o elemento anafórico e o acesso ao seu respectivo antecedente, maior a carga informacional desse elemento. Por isso, nomes repetidos seriam menos eficientes, pois têm mais traços semânticos a serem processados para identificar o seu respectivo antecedente, do que os pronomes. Além disso, um elemento anafórico precisa estar justificado em termos de sua função discursiva dentro de um contexto específico.

Em resumo, o autor explica a Penalidade do Nome Repetido argumentando que processar um nome repetido correferencial implica em acessar mais material fonológico e em reativar mais informação do que o necessário para recuperar um antecedente que já está em foco, e que o pronome tem menos informações para acessar na hora do processamento, por isso gera menos custo operacional da memória de trabalho.

Além desta explicação relacionada à Penalidade do Nome Repetido, comparando retomadas com pronomes e nomes repetidos, Almor (1999) utilizou também as retomadas anafóricas estabelecendo relação de hiperonímia com seu respectivo antecedente. Assim a forma de SNs mais específicos (ex: *pardal*) são processadas mais lentamente do que SNs mais gerais (ex: *animal*) em relação a um mesmo antecedente (ex: *pássaro*). Ou seja, (b) foi lido mais rapidamente do que (a).

- (a) Foi o *pássaro* que comeu a fruta. O *pardal* parecia bem satisfeito.
- (b) Foi o *pássaro* que comeu a fruta. O *animal* parecia bem satisfeito.

A partir desse resultado, concluiu-se que as retomadas com SNs mais específicos além de conterem mais traços que possibilitam a correferência com o seu antecedente, trazem adicionalmente informação nova, tornando a carga informacional maior, gerando assim um custo maior de tempo no processamento correferencial.

Para comprovar a hipótese relacionada ao custo operacional da memória de trabalho, Almor et al. (1999) realiza um estudo com pacientes com síndrome de Alzheimer. Ele compara os resultados desses pacientes, que contêm problemas de memória de trabalho, com os resultados obtidos com o grupo controle que não tinha nenhuma patologia.

Os resultados mostraram que os pacientes com a doença de Alzheimer tinham maior custo de processamento no estabelecimento da correferência quando as retomadas anafóricas eram feitas com pronomes do que quando eram feitas através de nomes repetidos. Este resultado é exatamente inverso ao encontrado para o grupo controle, em que o estabelecimento da correferência era facilitado com o uso de pronomes. Isso mostra que realmente há relação entre a memória de trabalho e custo de processamento. Quando o paciente tinha déficit de memória de trabalho, a correferência foi facilitada utilizando um nome repetido que tem mais informações para recuperar seu antecedente do que um pronome. Podemos dizer que com o uso de um nome repetido houve uma reativação das informações do seu antecedente, facilitando assim a sua ativação na memória de trabalho. Em outras palavras, embora a fala de pacientes com Alzheimer seja caracterizada pelo uso frequente e inapropriado de pronomes, a habilidade de compreendê-los está comprometida devido ao decréscimo geral na ativação do referente na memória de trabalho, sendo o uso de nomes repetidos mais funcional neste caso; o contrário do que acontece com indivíduos sem patologia.

A Hipótese da Carga Informacional abrangeu a explicação sobre a Penalidade do Nome Repetido, inicialmente proposta pela Teoria da Centralização, pois mostra resultados que mostram que não só entre retomadas pronominais e retomadas com SNs repetidos ocorre distinção no processamento correferencial, mas também com os diferentes tipos de SNs anafóricos. Ou seja, o que faz uma anáfora ser mais ou menos aceita em qualquer contexto não é a sua classe formal (pronome ou nome repetido), mas sim a relação entre seu custo de processamento e a função que esta anáfora exerce no discurso. Assim, a penalidade não deveria ser restrita apenas às retomadas anafóricas com nome repetido, mas deveria estender-se também às relações de processamento entre SNs mais específicos e SNs mais gerais.

Como vimos, há vários estudos psicolinguísticos que mostram que a hierarquia do elemento anafórico na sentença influencia diretamente no processamento do seu elemento correferencial. Entretanto, nos estudos apresentados até aqui só havia experimentos com sentenças que continham elementos anafóricos considerados proeminentes no discurso, sendo sujeito gramatical ou foco do discurso.

Para entendermos melhor sobre esse efeito, a seguir faremos uma breve introdução sobre os conceitos de foco e sua relação com o processamento correferencial. Serão expostos os principais tipos de foco, em seguida, iremos especificar e justificar o conceito adotado no presente estudo. Na sequência, apresentaremos os principais estudos experimentais sobre foco na área da psicolinguística com enfoque no processamento linguístico.

5) Conceitos de foco

O foco é um recurso empregado pelo falante para fazer com que o ouvinte fique atento a uma parte do enunciado, que pode ser um constituinte lexical, ou parte de uma sentença, ou uma entonação diferenciada. Podemos interpretá-lo de diversas formas. Há quem considere que a marcação de foco deva se refletir na estrutura sintática das frases e, assim, seja dependente de operações sintáticas abstratas (Cinque, 1993; Rizzi, 1997; Ordóñez, 1997; Costa, 1998). Outros consideram que o foco é formado através de

elementos de estrutura informacional (como informações contextuais do tipo dado/novo (Givon, 1995)) (Zubizarreta, 1998; Kiss, 1998). Para outros, o fenômeno pode ser prosodicamente motivado (Zubizarreta, 1998; Miotto, 2006).

Zubizarreta (1998) considera que quando o foco não é explicitado pela sintaxe, temos que recorrer ao contexto para identificá-lo. Neste caso, a autora diz que há dois tipos de foco que são diferenciados segundo o traço semântico de contraste. É chamado de foco não contrastivo a informação que é dada para completar ou responder a um contexto prévio. Ou seja, é fixada uma interrogativa-Wh como um contexto prévio. O foco não contrastivo pode ser considerado também como a informação que a pergunta solicita. Vejamos o exemplo a seguir, retirado de Quarezemin (2012):

- (1) a. O que João comeu?
- b. O João comeu [a torta].

O constituinte [a torta] responde à pergunta (1a) substituindo a expressão –Wh, portanto é o foco da sentença (1b). Já o foco contrastivo é identificado como a informação que funciona como uma correção, ou seja, apresenta um contraste em relação a outro constituinte. Vejamos o exemplo a seguir, retirado de Quarezemin (2012):

- (2) a. O João comeu um pastel.
- b. O João comeu UMA TORTA. (não um pastel)

Este foco nega o valor da variável da afirmação contextual para atribuir um novo valor a ela. Ou seja, em (2b) o falante nega o valor estabelecido pela variável [um pastel] e atribui um novo valor a ela [uma torta].

Já Kiss (1998) considera que há dois tipos de foco que podem variar segundo a variação dos traços semânticos [+exaustivo] ou [-exaustivo]. Quando o valor deste traço é positivo, temos o foco de identificação e a leitura deve ser [x e apenas x], como vemos no exemplo a seguir retirado de Quarezemin (2012):

- (1) Foi UM CARRO que a Maria ganhou.

Neste caso há um deslocamento do foco da posição de objeto indicando uma identificação exaustiva. Este tipo de foco também é chamado pela autora de foco de identificação. Já o foco com traço [-exaustivo] que introduz uma informação nova sem nenhum contraste, chama-se foco de informação.

Observamos que o que diferencia a proposta de Kiss (1998) e Zubizarreta (1998) é o foco de identificação, que não é considerado por Zubizarreta.

Existe também a articulação foco-suposição que é representada pela configuração sintática (Rizzi, 1997) e interpretada pelo componente de interface LF (*Logical Form*). A suposição é a informação que é considerada o plano de fundo da sentença e que supõe ser partilhada pelos falantes. Chierchia (2003) afirma que A

pressupõe B, e B é dada como certa, em todo contexto no qual A é usada. Observemos o exemplo a seguir:

- (4) a. Foi o João que parou de fumar.
b. Alguém parou de fumar.

Em (4) temos a sentença “a” que é uma sentença clivada que pressupõe “b”; esta tem que ser dada como verdadeira quando aquela é usada. Então, podemos concluir que a verdade de (4b) é tratada como não controversa quando (4a) é verdadeiro. No caso de não existir alguém que tenha parado de fumar, o uso de (4a) é completamente inaceitável.

Como vimos até aqui, o contexto tem um papel muito importante para a identificação do foco no discurso. Entretanto, existe uma forma explícita de configurar o foco na sentença, ou no discurso local. Pode-se recorrer à sintaxe, através das sentenças clivadas, para determinar o foco de maneira explícita sem a necessidade de introduzir um contexto.

A clivagem é feita com auxílio de itens designados especificamente para este fim, ou seja, os itens preparam um lugar específico para acomodar o constituinte eleito como foco. Em outras palavras, o constituinte, ou palavra focalizada, fica situado entre a cópula *foi* e o complementizador *que*, configurando-se sintaticamente desta forma: ser + XP + *que* + IP. Essa cópula determina o foco sintaticamente sem que precisemos recorrer a um contexto preestabelecido. De fato, se aplicarmos o processo de clivagem à sentença (5a), obteremos (5b), com o sujeito gramatical na posição de foco. Assim como podemos modificar o foco da sentença se movermos outro constituinte para a cópula (5c).

- (5) a. João parou de fumar.
b. Foi João que parou de fumar.
c. Foi de fumar que João parou.

Sendo assim, podemos concluir que há um foco estrutural, definido sintaticamente, nas sentenças (5b) e (5c). Por ser estabelecido sintaticamente, esse tipo de foco não necessita de um contexto prévio para ser definido como tal.

Na seção seguinte, iremos apresentar alguns estudos que utilizaram esse tipo de estrutura, clivada, para investigarem o efeito do foco no processamento da correferência.

5.1) Trabalhos sobre foco em processamento linguístico

Há poucos estudos na área de psicolinguística experimental que investigam o efeito de foco estrutural, principalmente estudos direcionados ao processamento correferencial.

Entre os principais estudos que investigam esse tema podemos destacar o de Almor & Eimas (2008) que, por meio de dois experimentos, investigou como o foco sintático afetaria o processamento de anáforas sintagmas nominais (SN) repetidos ou

pronomes, na linguagem falada. Para isso, os autores aplicaram dois experimentos. O primeiro experimento foi delineado para avaliar o efeito que o foco estrutural causa sobre o tratamento da anáfora (SN) repetida, na compreensão da linguagem falada, medindo latências de decisão lexical e evocação tardia para essas anáforas.

A hipótese na tarefa de decisão lexical era que decisões lexicais mais rápidas, com o antecedente em posição de foco estrutural, indicariam que o foco facilita o processamento, apesar do nome repetido. Decisões lexicais mais lentas, quando o antecedente estiver em posição de foco estrutural, indicariam que a penalidade do nome repetido ocorre cedo no processamento. E na tarefa de “*recall*”, evocação tardia, a penalidade do nome repetido afetaria o desempenho quando o antecedente fosse o foco da sentença. Assim, *pássaro* é acessado mais rapidamente em (1a) do que em (1b) (Almor & Eimas, 2008):

(1a) Foi o pássaro que comeu a fruta. O *pássaro* parecia bem satisfeito.

(1b) O que o pássaro comeu foi a fruta. O *pássaro* parecia bem satisfeito.

Os resultados desse primeiro experimento mostram que as decisões lexicais para anáforas com antecedentes na posição de foco foram significativamente mais rápidas do que quando não estavam em foco. E na tarefa de “*recall*” para anáforas com antecedentes na posição de foco foi significativamente pior quando retomadas por um SN repetido. Embora os tempos de decisão lexical tenham mostrado uma facilitação no efeito do foco, os resultados do “*recall*” demonstraram que há uma interferência do antecedente na representação da memória. Isso indica um efeito similar ao da Penalidade do Nome Repetido, na linguagem falada, nesse tipo de tarefa.

O segundo experimento tem a mesma metodologia que o Experimento 1, mas os autores utilizaram itens em que não havia anáforas SN repetidas, e mantiveram antecedente em foco ou não focalizado, sendo este antecedente um termo mais geral do que o termo correferente. Assim, *pássaro* é lido mais rapidamente em (2a) do que em (2b) (Almor & Eimas, 2008):

(2a) Foi o *pardal* que comeu a fruta. O *pássaro* parecia bem satisfeito.

(2b) O que o pardal comeu foi a *fruta*. O *pássaro* parecia bem satisfeito.

Os resultados deste experimento mostraram efeito facilitador do foco somente na tarefa de decisão lexical. A partir dos resultados destes experimentos podemos concluir que ao nos referirmos ao elemento anafórico mais ativado no discurso (em foco), falantes e ouvintes em geral usam uma forma anafórica com menor carga informacional. Outros experimentos de compreensão mostram que as anáforas que são mais gerais do que seus antecedentes, como os hiperônimos, são acessadas mais rapidamente do que quando são mais específicas, como os hipônimos (Leitão, 2005). Por outro lado, quando a anáfora é um nome repetido, é lida mais lentamente quando o referente está focalizado (Almor, 1999).

De acordo com o ponto de vista de Almor (1999), pronomes são processados mais rapidamente do que nomes repetidos no estabelecimento da correferência. Ele

explica esta afirmação em função da carga informacional que o nome repetido contém. Ou seja, quanto maior a distância semântica entre o elemento anafórico e o acesso ao antecedente, maior a carga informacional desse elemento. Por isso, nomes repetidos seriam menos eficientes, pois têm mais traços semânticos a serem processados para identificar o seu respectivo antecedente, do que os pronomes, o que torna o processamento mais custoso em termos de memória de trabalho.

Em resumo, os resultados obtidos por Almor & Eimas (2008) mostram que, semelhante à leitura, como em Almor (1999), que será abordado posteriormente, a compreensão da linguagem verbal é dificultada pela correferência com um nome repetido quando um referente está focalizado no discurso. E que anáforas com termos mais gerais do que seus antecedentes podem facilitar o processamento.

O estudo de Almor & Eimas (2008) citado anteriormente, sobre a linguagem falada, tomou como base o artigo de Almor (1999), que investigou praticamente as mesmas questões, só que na linguagem escrita, procurando justificar o processamento de diversas formas anafóricas segundo a sua função semântica no discurso. Foram investigadas anáforas do tipo SN repetido, hiperônimos, hipônimos e pronomes, manipulando a focalização dos seus respectivos antecedentes. Foram cinco experimentos de leitura automonitorada. O primeiro experimento procurou medir o efeito de foco no processamento de anáfora SN que não acrescentava novas informações sobre o seu referente e que tinha baixa carga informativa em relação ao seu antecedente. Ou seja, foi utilizado correferente anafórico que estabelecia uma relação de hiperonímia com seu antecedente que, por sua vez, poderia estar focalizado ou não.

Os resultados do primeiro experimento mostram que quando a anáfora era um termo mais geral, um hiperônimo, tendo assim uma carga informativa baixa em relação ao seu antecedente, e não acrescentando novas informações, o foco facilitou o processamento. Estes resultados confirmam a primeira previsão da Hipótese da Carga Informacional.

O experimento 2 foi realizado para testar a segunda previsão da Hipótese da Carga Informacional, ou seja, que a anáfora que tem uma carga informativa alta, mas adiciona novas informações sobre o seu antecedente, é mais fácil de processar quando seu antecedente é o foco da sentença. Ou seja, as anáforas que são mais específicas do que os seus antecedentes, em relação de hiponímia, têm uma carga informativa alta, mas também adicionam novas informações sobre o seu referente. Portanto, com a segunda previsão da Hipótese da Carga Informacional, as anáforas devem ser lidas mais rapidamente quando o seu antecedente estiver focalizado. Os resultados mostraram efeito significativo, com leitura mais rápida, quando o antecedente estava focalizado do que quando não estava, confirmando a segunda previsão da Hipótese da Carga Informacional. Ou seja, quando o correferente tem alta carga informacional, mas tem uma função de acrescentar informação nova, o processamento é mais rápido quando o antecedente está focalizado.

Já o terceiro experimento procurou testar a terceira previsão da Hipótese da Carga Informacional, que diz que quando uma anáfora com uma carga informativa alta não adiciona nenhuma informação nova, o processamento só é facilitado quando o antecedente não estiver focalizado. Para isso, o autor utilizou nomes repetidos como

anáforas de antecedentes, focalizados ou não. Os resultados mostram que anáforas repetidas foram lidas mais lentamente quando seu antecedente estava em foco. Este resultado além de apoiar a terceira predição da Hipótese da Carga Informacional, é compatível com resultados de estudos anteriores, que demonstraram a Penalidade do Nome Repetido (Gordon et al., 1993). De fato, os resultados do terceiro experimento estendem esta descoberta anterior, mostrando que a penalidade do nome repetido também é constatada pela diferença entre os tempos de leitura, sob diferentes condições de focalização.

O quarto experimento foi elaborado para testar se os resultados do terceiro foram influenciados pela configuração referencial. Ou seja, o foco é apenas informativo e não há um contraste entre o elemento focalizado e a repetição do mesmo. No quarto experimento foi feito um contraste utilizando dois elementos anafóricos possíveis, focalizando-se um desses elementos, em seguida, fazendo a correferência com o nome repetido ou com outro elemento não focalizado. Os resultados convergiram com os resultados obtidos no terceiro experimento, ou seja, houve Penalidade do Nome Repetido, também quando havia um contexto de contraste.

Outro estudo sobre foco que investiga questões de processamento anafórico é o de Cowles (2005), que procura testar a Hipótese da Carga Informacional manipulando o tipo de antecedente. Ou seja, o autor procura testar as mesmas questões abordadas em Almor (1999), relacionadas à hierarquia de acessibilidade de anáforas que tenham relação de hponímia ou hiperonímia com seus antecedente, e anáforas com nome repetido. Ele questiona os resultados obtidos pela configuração do antecedente possível, focalizado através de clivagem. Os resultados obtidos nesse estudo confirmam os resultados da hierarquia funcional das anáforas postulados pela Hipótese da Carga Informacional, e acrescentam que tal hierarquia não está restrita apenas a sentenças com antecedentes focalizados através de clivagem, mas também a sentenças com antecedente em posição de sujeito gramatical.

Em francês, o estudo de Fossard (1999) também mostrou efeito facilitador do foco estrutural no processamento da correferência em relação ao pronome pleno comparado com o nome repetido.

Vimos que as pesquisas feitas até aqui sobre o efeito do foco, através das técnicas experimentais, sempre procuram investigar as questões da Hipótese da Carga Informacional, e a Penalidade do Nome Repetido. Outro fator que pode influenciar no processamento da correferência é o paralelismo estrutural. A seguir, faremos um apanhado geral dos estudos que investigaram tal fator.

6) Paralelismo estrutural

Ocorre o paralelismo estrutural quando um elemento correferencial tem a mesma função e posição sintática que seu antecedente. Esse fator prevê que o processamento de pronome é facilitado quando está em paralelismo estrutural com seu antecedente. Ou seja, um antecedente na posição de sujeito será retomado por um pronome de maneira mais natural se este estiver também na posição de sujeito. O mesmo aconteceria para um antecedente na posição de objeto, isto é, a retomada anafórica pronominal seria facilitada quando acontecesse também paralelamente em posição de objeto.

Estudos preliminares como os de Grober et al. (1978) e Garvey, Caramazza & Yates (1976) mostram que o paralelismo estrutural é um fator relevante, sendo uma estratégia robusta na resolução anafórica de pronomes. Entretanto, estes autores não levaram em consideração a interpretação de pronomes na posição de objeto.

Estudos posteriores (Smyth, 1992 e 1994) mostram que quando há congruência dos papéis temáticos, o efeito do paralelismo também é realçado, ou seja, o processamento se torna mais rápido. Se esse critério de congruência não existe, então o efeito facilitador do paralelismo é consideravelmente reduzido (Stevenson, Nelson & Stenning, 1995; Smyth, 1994).

Um estudo posterior (Chambers & Smyth, 1998) tem como principal objetivo testar alguns conceitos da Teoria da Centralização. Tais como proeminência sintática e a Penalidade do Nome Repetido para o único *Cb* que retomam um correferente em posição de sujeito. Para isso, os autores utilizaram técnicas de experimentos *on-line* e *off-line* variando em grupos de sentenças com retomada anafórica, tanto em posição de sujeito quanto na posição de objeto, além de sentenças com paralelismo e outras sem.

No primeiro experimento, o principal objetivo foi testar o efeito de paralelismo como sendo um fator importante no processamento da correferência, diferentemente dos outros estudos que tratavam esse efeito como um fator secundário, assim como testar tal efeito em posição de objeto. Foi usada a técnica *off-line* de julgamento. Os resultados do primeiro experimento mostraram que há uma preferência pelas estruturas paralelas.

No experimento 2, o objetivo, assim como no primeiro experimento, foi observar o efeito de paralelismo e ainda testar o conceito da Teoria da Centralização referente à Penalidade do Nome Repetido, aplicando à retomada anafórica em posição de objeto. Desta vez foi utilizada a técnica *on-line* de leitura automonitorada. Foram manipuladas frases com antecedente paralelo (em posição de sujeito e em posição de objeto), e frases com antecedente não paralelo (em posição de sujeito e em posição de objeto), além do preâmbulo e da sentença final. Os resultados mostraram que a leitura das sentenças com pronome, seja em posição de sujeito ou de objeto, em retomada com o antecedente paralelo, foram lidas significativamente mais rápido do que as com antecedente em posição não paralela. O que contraria o conceito de proeminência sintática baseada no *Cf* na posição de sujeito gramatical como sendo a posição proeminente sintaticamente.

Em resumo, a questão levantada por Chambers & Smyth (1998) foi que a Penalidade do Nome Repetido, que ficou constatada na retomada com antecedentes paralelos, seja em posição de sujeito ou de objeto, contraria a Teoria da Centralização que prevê apenas penalidade para o único centro anafórico *Cb* na posição de sujeito.

7) Conclusão

A correferência pode ser estabelecida de variadas formas, através de pronome, nome repetido, categoria vazia, ou mesmo por meio de SNs que estabelecem uma relação de hiperonímia ou hiponímia com seu antecedente. Essas formas linguísticas correferenciais podem aparecer em posição de sujeito ou de objeto, tanto no escopo da sentença (intrassentencial), quanto no escopo do discurso (intersentencial). O objetivo principal de se observar o processamento de tais formas referenciais, é buscar a

compreensão dos fatores semânticos e estruturais que atuam no processo da correferência e integração discursiva, e observar de que maneira ocorre essa atuação.

Os estudos que observam as relações anafóricas intersentenciais procuram observar como fatores tanto estruturais e gramaticais (por exemplo: paralelismo estrutural, concordância), quanto fatores semântico-pragmáticos (por exemplo: tipo de retomada, saliência discursiva ou proeminência sintática) podem influenciar no estabelecimento da correferência.

Os estudos apresentados neste artigo colocaram em evidência que a utilização de um pronome anafórico para designar uma entidade saliente - ou seja, em foco no discurso - traz um índice explícito, deixando-o mais ativo na memória de trabalho, enquanto que o emprego de uma expressão mais informativa, como um nome próprio, ou SN definido, torna essa relação mais difícil de ser estabelecida, gerando assim um custo de processamento denominado Penalidade do Nome Repetido.

Entretanto, no estudo de Chambers & Smyth (1998) ficou constatada a Penalidade do Nome Repetido, seja em posição de sujeito ou de objeto, quando retomada e antecedente estavam em função e posição de paralelismo. Esses resultados mostram que o paralelismo é um fator importante no estabelecimento da correferência.

A Hipótese da Carga Informacional propõe outra explicação para este fenômeno psicolinguístico. A Hipótese relaciona o processamento correferencial com a distância semântica entre os elementos anafóricos e seus respectivos antecedentes e o custo operacional da memória de trabalho. Os pronomes são processados mais rapidamente do que nomes repetidos no estabelecimento da correferência, em função da carga informacional que essa forma linguística contém, ou seja, quanto maior a distância semântica entre o elemento anafórico e o acesso ao seu respectivo antecedente, maior a carga informacional desse elemento. Por isso, nomes repetidos seriam menos eficientes, pois têm mais traços semânticos a serem processados para identificar o seu respectivo antecedente, do que os pronomes que evocam uma representação mais geral.

Outro fator que pode atuar no processamento correferencial é o foco. O foco pode ser determinado tanto pela prosódia quanto pela estrutura sintática, ou por partes de uma frase ou enunciados que podem contribuir para a saliência de uma determinada informação. Mostramos nos estudos apresentados anteriormente que o foco age como um fator facilitador no estabelecimento da correferência.

Todos estes estudos relatados acima nos levam a questionar se há fatores que atuam mais fortemente facilitando o estabelecimento da correferência do que outros, ou se um fator pode mascarar o efeito de outro quando confrontados no mesmo estímulo experimental. Ou melhor, se há uma hierarquia de fatores que atuam facilitando a acessibilidade na hora do processamento anafórico e, se realmente houver, como se dá essa hierarquia.

8) Referências bibliográficas

ALMOR, A. Noun-phrase anaphora and focus: the informational load hypothesis. *Psychological Review*, 106, p. 748-765, 1999.

_____. Constraints and mechanisms in theories of anaphor processing. In: Pickering et alii. (Ed.). **Architectures and mechanisms for language processing**. Cambridge University Press. England, 2000.

ALMOR A. & EIMAS, P. D. Focus and noun phrase anaphors in spoken language comprehension, **Language and Cognitive Processes**, 23 (2), p. 201-225, 2008.

CHAMBERS, C. & SMYTH, R. Structural parallelism and discourse coherence: a test of centering theory. **Journal of Memory and Language**. 39, p. 593-608, 1998.

CHOMSKY, N. **Lectures on government and binding**. Dodrecht: Foris 1981.

COWLES, H. W., & GARNHAM, A. Antecedent focus and conceptual distance effects in category noun-phrase anaphora. **Language and Cognitive Processes**, 20, p.725-750, 2005.

ERMST, EMILIE. Le traitement en temps réel de l'anaphore pronominal le dans le langage écrit - Développement normal et dysfonctionnements. **Apports de la théorie du Centrage**. Paris: Université Paris V, 2007.

FOSSARD, M. Traitement anaphorique et structure du discours. Etude psycholinguistique des effets du focus de discours sur la spécificité de deux marqueurs référentiels: le pronom anaphorique "il" et le nom propre répété. In **Cognito**, 15, p.33-40, 1999.

GARVEY, C., CARAMAZZA, A., & YATES, J. Factors affecting assignment of pronoun antecedents. **Cognition**. 32, p.227-243, 1976.

GORDON, P. C., GROSZ, B. J., & GILLION, L. A. Pronouns, names, and the centering of attention in discourse. **Cognitive Science**. 17, p.311-347, 1993.

GORDON, P. C. & CHAN. Pronouns passives and discourse coherence. **Journal of Memory and Language**, 34, p. 216-231, 1995.

GORDON, P. C. & HENDRICK, R. The representation and processing of coreference in discourse. **Cognitive Psychology**, 22, p. 389-424, 1998.

GROBER, E. H., BEARDSLEY, W., & CARAMAZZA, A. Parallel function strategy in pronoun assignment. **Cognition**. 6, p. 117-133, 1978.

GROSS B. J., JOSHI, A. K., & WEINSTEIN, S. Centering: A framework for modelling the local coherence of discourse. **Computational Linguistics**. 21, p. 203-226, 1995.

KENNISSON, S., & GORDON, P. C. Comprehending referential expressions during reading: evidence from eye tracking. **Discourse Processes**. 24, p. 229-252, 1997.

KISS, K. Identificational focus versus information focus. **Language**, 74 (2), p. 245-273, 1998.

LEITÃO, M. M. **O processamento do objeto direto anafórico no Português Brasileiro**. Rio de Janeiro, UFRJ, Faculdade de Letras. 149 fls. Tese de Doutorado em Linguística, 2005.

MIOTO, C. Focalização e Quantificação. **Revista Letras**. Curitiba : Editora UFPR, 61, p. 169-189, 2006.

SMYTH, R. H. Multiple feature matching in pronoun resolution: a new look at parallel function. **Proceedings of the second international conference on spoken language processing**. Edmonton; Priority Printing. p. 145-148, 1992.

_____. Grammatical determinants of ambiguous pronoun resolution. **Journal of Psycholinguistic Research**. 23, p. 197-229, 1994.

YANG, C. L., GORDON, P. C., HENDRICK, R., WU, J. T., & CHOU, T. L. The processing of coreference for reduced expressions in discourse integration. **Journal of Psycholinguistic Research**. 30, p. 21-35, 2001.

ZUBIZARRETA, M. L. **Prosody, Focus and Word Order**. Cambridge, 1998 MA: MIT Press.